

# Testing for Granger causality in large mixed-frequency VARs

Citation for published version (APA):

Götz, T. B., Hecq, A., & Smeekes, S. (2016). Testing for Granger causality in large mixed-frequency VARs. *Journal of Econometrics*, 193(2), 418-432. <https://doi.org/10.1016/j.jeconom.2016.04.015>

## Document status and date:

Published: 01/08/2016

## DOI:

[10.1016/j.jeconom.2016.04.015](https://doi.org/10.1016/j.jeconom.2016.04.015)

## Document Version:

Accepted author manuscript (Peer reviewed / editorial board version)

## Document license:

CC BY-NC-ND

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

# TESTING FOR GRANGER CAUSALITY IN LARGE MIXED-FREQUENCY VARs

Thomas B. Götz<sup>\*1</sup>, Alain Hecq<sup>†2</sup>, and Stephan Smeekes<sup>‡2</sup>

<sup>1</sup>Deutsche Bundesbank

<sup>2</sup>Maastricht University

March 22, 2016

## Abstract

We analyze Granger causality testing in a mixed-frequency VAR, where the difference in sampling frequencies of the variables is large, implying parameter proliferation problems in case we attempt to estimate the model unrestrictedly. We propose several tests based on reduced rank restrictions, including bootstrap versions thereof to account for factor estimation uncertainty and improve the finite sample properties of the tests, and a Bayesian VAR extended to mixed frequencies. We compare these methods to a test based on an aggregated model, the *max*-test (Ghysels et al., 2016a) and an unrestricted VAR-based test (Ghysels et al., 2016b) using Monte Carlo simulations. An empirical application illustrates the techniques.

**JEL Codes:** C11, C12, C32

**JEL Keywords:** Granger Causality, Mixed Frequency VAR, Bayesian VAR, Reduced Rank Model, Bootstrap Test

---

<sup>\*</sup>Correspondence to: Thomas B. Götz, Deutsche Bundesbank, Macroeconomic Analysis and Projection Division, Wilhelm-Epstein-Strasse 14, 60431 Frankfurt am Main. Email: [thomas.goetz@bundesbank.de](mailto:thomas.goetz@bundesbank.de), Tel: +49-69-9566-6991, Fax: +49-69-9566-6734. The views expressed in this paper are ours and do not necessarily reflect the views of the Deutsche Bundesbank or its staff. Any errors or omissions are our responsibility.

<sup>†</sup>Maastricht University, School of Business and Economics, Department of Quantitative Economics, P.O. Box 616, 6200 MD Maastricht, The Netherlands. Email: [a.hecq@maastrichtuniversity.nl](mailto:a.hecq@maastrichtuniversity.nl)

<sup>‡</sup>Maastricht University, School of Business and Economics, Department of Quantitative Economics, P.O. Box 616, 6200 MD Maastricht, The Netherlands. Email: [s.smeekes@maastrichtuniversity.nl](mailto:s.smeekes@maastrichtuniversity.nl)

# 1 Introduction

Economic time series are published at various frequencies. While higher frequency variables used to be aggregated (e.g., Silvestrini and Veredas, 2008), it has become more and more popular to consider models that take into account the difference in frequencies of the processes under consideration. As argued extensively in the mixed-frequency literature (among others, Ghysels et al., 2007, Clements and Galvão, 2008, Ghysels et al., 2006, Marcellino and Schumacher, 2010), working in a mixed-frequency setup instead of a common low-frequency one is advantageous due to the potential loss of information in the latter scenario and feasibility of the former through MI(xed) DA(ta) S(ampling) regressions (Ghysels et al., 2004), even in the presence of many high-frequency variables compared to the number of observations.

Until recently, the MIDAS literature was limited to the single-equation framework, in which one of the low-frequency variables is chosen as the dependent variable and the high-frequency ones are in the regressors. Since the work of Ghysels (2016) for stationary series and the extension of Götz et al. (2013) and Ghysels and Miller (2015) for the non-stationary and possibly cointegrated case, we can analyze the link between high- and low-frequency series in a VAR system treating all variables as endogenous. Note that the VAR vector in the model of Ghysels (2016) is specified in terms of the low frequency by stacking the high-frequency variables and aligning them with the low-frequency ones. Another approach of incorporating mixed frequencies into VAR models is to write the entire model in high frequency by treating the corresponding observations of the low-frequency series as latent. Noteworthy contributions to this branch of the mixed-frequency VAR literature are, among others, Schorfheide and Song (2015), Kuzin et al. (2011) or Eraker et al. (2015). We refer the reader to Forni et al. (2013) for an excellent survey on VAR models with mixed-frequency data.

An often investigated feature within VAR models is Granger causality, generally introduced in Granger (1969). It is well known, though, that Granger causality is not invariant to temporal

aggregation (Granger, 1988, Granger and Lin, 1995 or Marcellino, 1999). While most work on this issue focused on cases in which all series are aggregated, Ghysels et al. (2016b) investigate Granger causality in a setting, where variables are available at varying frequencies. Starting from an underlying high-frequency process, it is shown that mixed-frequency Granger causality tests better recover causal patterns in the data generating process than traditional low-frequency tests do. However, decent size and power properties of their mixed-frequency tests depend on a relatively small difference in sampling frequencies of the variables involved. Indeed, if the number of high-frequency observations within a low-frequency period is large, size distortions and loss of power may be expected.

In this paper we take the model of Ghysels (2016) and analyze the finite sample behavior of Granger non-causality tests in the spirit of Dufour and Renault (1998) when the number of high-frequency observations per low-frequency period is large as, e.g., in a month/working day-example. To avoid the proliferation of parameters we consider two parameter reduction techniques: reduced rank regressions and a Bayesian mixed-frequency VAR. As far as the latter is concerned, we show precisely how to extend the dummy observation approach of Bańbura et al. (2010), itself built on the work of Sims and Zha (1998), to the presence of mixed frequencies.<sup>1</sup> Importantly, due to stacking the high-frequency variables in the mixed-frequency VAR (Ghysels, 2016), their approach cannot be applied directly such that a properly adapted choice of auxiliary dummy variables is required, which marks a significant contribution to the respective literature by itself. With respect to reduced rank regressions, the associated factors are typically not observable and must be estimated, which obviously affects the distribution of the Wald tests to detect directions of Granger causalities. Consequently, another important contribution of this paper is the introduction of bootstrap versions of these tests (also for the unrestricted VAR), which have correct size even for large VARs and a small sample size.

---

<sup>1</sup>Bańbura et al. (2010) refer to these variables as dummy observations. To avoid confusion between high- and low-frequency *observations* and auxiliary *variables*, we use the term 'auxiliary dummy variables' henceforth.

Independently and simultaneously to this work, Ghysels et al. (2016a) introduced another approach to testing for Granger causality in a mixed-frequency VAR characterized by a large discrepancy in frequencies of the series involved. In particular, their method is based on a sequence of parsimonious regression models where the low-frequency variable is regressed on its own lags and only one specific lead or lag of the high-frequency series. The test statistic itself is then computed as the maximum of these individual-regression-based estimators scaled and weighted properly; hence the name “*max*-test”. The authors show that, in a mixed-frequency VAR à la Ghysels (2016), this procedure leads to correct identification of the test in the presence or absence of Granger causality, which implies consistency against causality of all forms. Its simplicity and applicability to large dimensional systems makes the *max*-test the main competitor for our tests.

Both reduced rank regressions and the Bayesian estimation of the mixed-frequency VAR, are aiming at reducing the amount of parameters to be estimated before testing for Granger causality. The former does so by imposing a certain set of restrictions leading to the computation of a low-dimensional set of factors; the latter uses parameter shrinkage to a prior that is centered at a restricted version of the VAR of Ghysels (2016). While these methods are very appealing from an applied perspective, they are, much like the MIDAS polynomials, entirely ad hoc and possibly misspecified. Indeed, if the restrictions imposed are not satisfied in the true data generating process, i.e., unless we impose (over)identified restrictions, the resulting model is not correctly specified. This, in turn, may result in potential inability of our methods to detect causality, and in particular cause our methods to lose consistency against certain forms of causality that cannot be detected under the misspecified restrictions. In contrast, Granger causality tests based on the unrestricted VAR (Ghysels et al., 2016b) and the *max*-test (Ghysels et al., 2016a) are not based on (possibly invalid) restrictions implying that they have asymptotic power of one against all forms of causality. However, in situations where our tests are based on a correctly specified model, our approaches may very well outperform the test based on

the unrestricted VAR or the *max*-test due to the way in which the parameter proliferation problem in our large dimensional system is addressed. Moreover, even in situations in which the restrictions are not valid, in finite samples our methods may still allow one to pick up causality with high power due to the dimensionality reduction which may offset the effect of misspecification in small samples.

The rest of the paper is organized as follows. In Section 2 notations are introduced, the mixed-frequency VAR (MF-VAR hereafter) for our specific case at hand is presented and Granger (non-)causality is defined. Section 3 discusses the approaches to reduce the number of parameters to be estimated, whereby reduced rank restrictions (Section 3.1) as well as Bayesian MF-VARs (Section 3.2) are presented in detail. The finite sample performance of testing for Granger causality within the aforementioned restricted models is analyzed via a Monte Carlo experiment in Section 4. We compare our tests with the one based on the unrestricted VAR, the *max*-test, and an asymptotic Wald test based on a common low-frequency VAR obtained by temporally aggregating the high-frequency variable (Breitung and Swanson, 2002). An empirical example with U.S. data on the monthly industrial production index and daily volatility in Section 5 illustrates the results. Section 6 concludes.

## 2 Causality in a Mixed-frequency VAR

### 2.1 Notation

Let us start from a two variable mixed-frequency system, where  $y_t$ ,  $t = 1, \dots, T$ , is the low-frequency variable and  $x_{t-i/m}^{(m)}$  are the high-frequency variables with  $m$  high-frequency observations per low-frequency period  $t$ . Throughout this paper we assume  $m$  to be rather large as in a year/month- or month/working day-example. We also assume  $m$  to be constant rather than varying with  $t$ .<sup>2</sup> The value of  $i$  indicates the specific high-frequency observation under con-

---

<sup>2</sup>As long as  $m$  is deterministic, even time-varying frequency discrepancies do not pose a problem on a theoretical level. However, the assumption of constant  $m$  simplifies the notation greatly (Ghysels, 2016).

sideration, ranging from the beginning of each  $t$ -period  $(x_{t-(m-1)/m}^{(m)})$  until the end  $(x_t^{(m)})$  with  $i = 0$ ). These notational conventions have become standard in the mixed-frequency literature and are similar to the ones in Götz et al. (2014), Clements and Galvão (2008, 2009) or Miller (2014).

Furthermore, let  $\underline{W}_t = (W'_{t-1}, W'_{t-2}, \dots, W'_{t-p})'$  denote the last  $p$  low-frequency lags of any process  $W$  stacked. Finally,  $\mathbf{0}_{i \times j}$  ( $\mathbf{1}_{i \times j}$ ) denotes an  $(i \times j)$ -matrix of zeros (ones),  $I_i$  is an identity matrix of dimension  $i$ ,  $\otimes$  represents the Kronecker product and  $vec$  corresponds to the operator stacking the columns of a matrix.

**Remark 1** *Extensions towards representations of higher dimensional multivariate systems as in Ghysels et al. (2016b) can be considered, but are left for further research here. Analyzing Granger causality among more than two variables inherently leads to multi-horizon causality (see Lütkepohl, 1993 among others). The latter implies the potential presence of a causal chain: for example, in a trivariate system,  $X$  may cause  $Y$  through an auxiliary variable  $Z$ . To abstract from that scenario, Ghysels et al. (2016b) often consider cases in which high- and low-frequency variables are grouped and causality patterns between these groups, viewed as a bivariate system, are analyzed. They study the presence of a causal chain and multi-horizon causality in a Monte Carlo analysis though.*

## 2.2 MF-VARs

Considering each high-frequency variable such that  $X_t^{(m)} = (x_t^{(m)}, x_{t-1/m}^{(m)}, \dots, x_{t-(m-1)/m}^{(m)})'$ , a dynamic structural equations model for the stationary multivariate process  $Z_t = (y_t, X_t^{(m)'})'$  is given by  $A_c Z_t = c + A_1 Z_{t-1} + \dots + A_p Z_{t-p} + \varepsilon_t$ . Note that the parameters in  $A_c$  are related to the ones in  $A_1$  due to stacking the high-frequency observations  $X_t^{(m)}$  in  $Z_t$  (Ghysels, 2016).<sup>3</sup>

---

<sup>3</sup>Compared to Ghysels (2016) we simply reverse the mixed-frequency vector  $Z_t$  and put the low-frequency variable first.

Explicitly for a lag length of  $p = 1$ , the model reads as:

$$\begin{pmatrix} 1 & \beta_1 & \beta_2 & \dots & \beta_m \\ \delta_1 & 1 & -\rho_1 & \dots & -\rho_{m-1} \\ \delta_2 & 0 & 1 & \dots & -\rho_{m-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \delta_m & 0 & 0 & \dots & 1 \end{pmatrix} \begin{pmatrix} y_t \\ x_t^{(m)} \\ x_{t-1/m}^{(m)} \\ \vdots \\ x_{t-(m-1)/m}^{(m)} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_{m+1} \end{pmatrix} + \begin{pmatrix} \rho_y & \theta_1 & \theta_2 & \dots & \theta_m \\ \psi_2 & \rho_m & \dots & \dots & 0 \\ \psi_3 & \rho_{m-1} & \rho_m & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \psi_{m+1} & \rho_1 & \rho_2 & \dots & \rho_m \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1}^{(m)} \\ x_{t-1-1/m}^{(m)} \\ \vdots \\ x_{t-1-(m-1)/m}^{(m)} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \vdots \\ \varepsilon_{(m+1)t} \end{pmatrix}. \quad (1)$$

In general, we assume the high-frequency process to follow an  $\text{AR}(q)$  process with  $q < m$  (we set  $q = m$  in the equation above; if  $q < m$ , one should set  $\rho_{q+1} = \dots = \rho_m = 0$ ). For non-zero values of  $\beta_j$  or  $\delta_j$  the matrix  $A_c$  links contemporaneous values of  $y$  and  $x$ , a feature referred to as nowcasting causality (Götz and Hecq, 2014).<sup>4</sup>

Pre-multiplying (1) by  $A_c^{-1}$  we get to the mixed-frequency reduced-form  $\text{VAR}(p)$  model:

$$\begin{aligned} Z_t &= \mu + \Gamma_1 Z_{t-1} + \dots + \Gamma_p Z_{t-p} + u_t \\ &= \mu + B' \underline{Z}_t + u_t. \end{aligned} \quad (2)$$

Consequently,  $\mu = A_c^{-1}c$ ,  $\Gamma_i = A_c^{-1}A_i$  and  $u_t = A_c^{-1}\varepsilon_t$ . Let  $B = (\Gamma_1, \Gamma_2, \dots, \Gamma_p)'$  and  $B(z) = I_{m+1} - \sum_{j=1}^p \Gamma_j z^j$ . We then make the following assumptions on the MF-VAR:

---

<sup>4</sup> $\beta_j \neq 0$  implies that  $y_t$  is affected by incoming observations of  $X_t^{(m)}$ , whereas  $\delta_j \neq 0$  implies that the high-frequency observations are influenced by  $y_t$  (see Götz and Hecq, 2014). The latter becomes interesting for studying policy analysis, where the high-frequency policy variable(s) may react to current low-frequency conditions (see Ghysels, 2016 for details).



**Assumption 2**  $Z_t$  is generated by the MF-VAR( $p$ ) in (2), for which it holds that (i) the roots of the matrix polynomial  $B(z)$  all lie outside the unit circle; (ii)  $u_t$  is independent and identically distributed (i.i.d.) with  $\mathbb{E}(u_t) = 0$ ,  $\mathbb{E}(u_t u_t') = \Sigma_u$ , with  $\Sigma_u$  positive definite, and  $\mathbb{E}(\|u_t\|^4) < \infty$ , where  $\|\cdot\|$  is the Frobenius norm.

Assumption 2(i) ensures that the MF-VAR is I(0), while (ii) is a standard assumption to ensure validity of the bootstrap for VAR models, see, e.g., Paparoditis (1996), Kilian (1998) or Cavaliere et al. (2012).

**Remark 3** The i.i.d. assumption on  $u_t$  is not strictly necessary to derive the limit distributions of the test statistics and could be weakened to a martingale difference sequence (mds) assumption (see, e.g., Ghysels et al., 2016b). However, it allows us to directly use the aforementioned literature on the validity of our bootstrap method that uses i.i.d. resampling of the VAR residuals, as considered in the simulation study in Section 4. Asymptotic validity of the bootstrap can also be established for heteroskedastic error terms, provided one adapts the bootstrap method to a wild bootstrap as we do in the empirical application in Section 5. Asymptotic validity of the wild bootstrap for univariate autoregressions is established in Gonçalves and Kilian (2004) and extended to VAR models by Brüggemann et al. (2016). Section 3.1.4 provides further details on our bootstrap methods and their validity.

**Remark 4** Assumption 2 implies that the data are truly generated at mixed frequencies. Hence, similar to Ghysels et al. (2016a) but unlike Ghysels et al. (2016b), we do not start with a common high-frequency data generating process (DGP hereafter).<sup>5</sup> The latter would imply causality patterns to arise at the high frequency. Consequently, we do not investigate which causal relationships at high frequency get preserved when moving to the mixed- or low-frequency case.

---

<sup>5</sup>Given such a situation it would seem natural to cast the model in state space form and estimate the parameters using the Kalman filter. However, this amounts to a model containing latent processes, i.e., the high-frequency observations of  $y$ . The latter is a feature we try to avoid in our MF-VAR: say we are interested in the impact of shocks to one or several variables on the whole system. Using a high-frequency DGP with missing observations implies that shocks to these latent processes are also latent and unobservable. This is undesirable given that, e.g., policy shocks are, of course, observable (Forni and Marcellino, 2014).

Indeed, an extension of our methods along these lines would demand a careful analysis of the mixed- and low-frequency systems corresponding to their latent high-frequency counterpart (see Ghysels et al., 2016b for the unrestricted VAR case).

Equation (2) is easy to estimate for small  $m$ , yet becomes a rather large system as the latter grows. For example, in a MF-VAR(1),

$$\begin{pmatrix} y_t \\ x_t^{(m)} \\ x_{t-1/m}^{(m)} \\ \vdots \\ x_{t-(m-1)/m}^{(m)} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_{m+1} \end{pmatrix} + \underbrace{\begin{pmatrix} \gamma_{1,1} & \gamma_{1,2} & \cdots & \gamma_{1,m+1} \\ \gamma_{2,1} & \gamma_{2,2} & \cdots & \gamma_{2,m+1} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{m+1,1} & \gamma_{m+1,2} & \cdots & \gamma_{m+1,m+1} \end{pmatrix}}_{\Gamma_1} \quad (3)$$

$$\times \begin{pmatrix} y_{t-1} \\ x_{t-1}^{(m)} \\ x_{t-1-1/m}^{(m)} \\ \vdots \\ x_{t-1-(m-1)/m}^{(m)} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \\ \vdots \\ u_{(m+1)t} \end{pmatrix},$$

$$u_t \stackrel{i.i.d.}{\sim} (\mathbf{0}_{((m+1) \times 1)}, \Sigma_u), \quad \Sigma_u = \begin{pmatrix} \sigma_{1,1} & \sigma_{1,2} & \cdots & \sigma_{1,m+1} \\ \sigma_{2,1} & \sigma_{2,2} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m+1,1} & \cdots & \cdots & \sigma_{m+1,m+1} \end{pmatrix}, \quad (4)$$

there are  $(m+1)^2$  parameters to estimate in the matrix  $\Gamma_1$ . Additional lags would further complicate the issue.

### 2.3 Granger Causality in MF-VARs

Let  $\Omega_t$  represent the information set available at moment  $t$  and let  $\Omega_t^W$  denote the corresponding set excluding information about the stochastic process  $W$ . With  $P[X_{t+h}^{(m)}|\Omega]$  being the best linear forecast of  $X_{t+h}^{(m)}$  based on  $\Omega$ , Granger non-causality is defined as follows (Dufour and Renault, 1998):

**Definition 5**  *$y$  does not Granger cause  $X^{(m)}$  if  $P[X_{t+1}^{(m)}|\Omega_t^y] = P[X_{t+1}^{(m)}|\Omega_t]$ . Similarly,  $X^{(m)}$  does not Granger cause  $y$  if  $P[y_{t+1}|\Omega_t^{X^{(m)}}] = P[y_{t+1}|\Omega_t]$ .*

In other words,  $y$  does not Granger cause  $X^{(m)}$  if past information of the low-frequency variable do not help in predicting current (or future) values of the high-frequency variables and vice versa. In terms of (3), testing for Granger non-causality implies the following null (and alternative) hypotheses:  $H_0 : \gamma_{2,1} = \dots = \gamma_{m+1,1} = 0$  ( $H_A : \gamma_{i,1} \neq 0$  for at least one  $i = 2, \dots, m+1$ ) for  $y$  not Granger causing  $X^{(m)}$  and  $H_0 : \gamma_{1,2} = \dots = \gamma_{1,m+1} = 0$  ( $H_A : \gamma_{1,i} \neq 0$  for at least one  $i = 2, \dots, m+1$ ) for  $X^{(m)}$  not Granger causing  $y$ .

## 3 Parameter Reduction

This section presents techniques that we have considered, and evaluated through a Monte Carlo exercise, with the aim to reduce the amount of parameters to be estimated in the MF-VAR model. Two approaches are discussed in detail, reduced rank restrictions and a Bayesian MF-VAR.

There are many alternative approaches to reduce the number of parameters among which are principal components, Lasso or ridge regressions. However, using principal components does not necessarily preserve the dynamics of the VAR under the null: nothing prevents, e.g., the first and only principal component to be loading exclusively on  $y$  implying that the remaining dynamics enter the error term. In other words, the autoregressive matrices in (2) may and will

most likely not be preserved, which naturally affects the block of parameters we test on for Granger non-causality. As for Lasso and ridge regressions, it is well known that they may be interpreted in a Bayesian context. In particular, the latter is equivalent to imposing a normally distributed prior with mean zero on the parameter vector (Vogel, 2002, among others), while the former may be replicated using a zero-mean Laplace prior distribution (Park and Casella, 2008). Given that our set of models contains a Bayesian VAR for mixed-frequency data, we abstract from its connection to regularized versions of least squares at this stage.

### 3.1 Reduced Rank Restrictions

#### 3.1.1 Reduced Rank Regression Model

In order to reduce the number of parameters to estimate in the MF-VAR, we propose the following reduce rank regression model, for which we make the following assumption:

**Assumption 6** *Let  $B_{X^{(m)}}$  be the matrix obtained from  $B'$  in (2) by excluding the first columns of  $\Gamma_1, \Gamma_2, \dots, \Gamma_p$ . The rank of this matrix is smaller than the number of high-frequency observations within  $Z_t$ , i.e.,  $rk(B'_{X^{(m)}}) = r < m$ .*

The model then reads as follows:

$$\begin{aligned} Z_t &= \mu + \gamma_{\cdot,1} \underline{y}_t + \alpha \sum_{i=1}^p \delta'_i X_{t-i}^{(m)} + \nu_t \\ &= \mu + \gamma_{\cdot,1} \underline{y}_t + \alpha \delta' \underline{X}_t^{(m)} + \nu_t, \end{aligned} \tag{5}$$

where  $\gamma_{\cdot,1}$  is the  $(m+1) \times p$  matrix containing the first columns of  $\Gamma_1, \Gamma_2, \dots, \Gamma_p$ , and  $\alpha$  and  $\delta = (\delta'_1, \dots, \delta'_p)'$  are  $(m+1) \times r$  and  $pm \times r$  matrices, respectively. Note that (5) can also be written in terms of  $\underline{Z}_t$ . Let us define  $\underline{\Gamma}_i \equiv (\gamma_{\cdot,1}^{(i)}, \alpha \delta'_i)$ , where  $\gamma_{\cdot,1}^{(i)}, i = 1, \dots, p$ , corresponds to the  $i^{th}$  column of  $\gamma_{\cdot,1}$ . Then, (5) is equivalent to  $Z_t = \mu + \underline{B}' \underline{Z}_t + \nu_t$ , where  $\underline{B} = (\underline{\Gamma}_1, \dots, \underline{\Gamma}_p)'$ .

For  $p = 1$  the model becomes

$$\begin{pmatrix} y_t \\ x^{(m)} \\ x_{t-1/m}^{(m)} \\ \vdots \\ x_{t-(m-1)/m}^{(m)} \end{pmatrix} = \mu + \underbrace{\begin{pmatrix} \gamma_{1,1} \\ \gamma_{2,1} \\ \vdots \\ \gamma_{m+1,1} \end{pmatrix}}_{\gamma_{\cdot,1}^{(1)}} y_{t-1} + \underbrace{\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{m+1} \end{pmatrix}}_{\alpha} \delta'_1 \begin{pmatrix} x_{t-1}^{(m)} \\ x_{t-1-1/m}^{(m)} \\ \vdots \\ x_{t-1-(m-1)/m}^{(m)} \end{pmatrix} + \nu_t, \quad (6)$$

where each  $\alpha_i, i = 1, \dots, m+1$ , is of dimension  $1 \times r$  and where  $\delta_1$  is an  $m \times r$  matrix. Hence, we could call  $\delta'_1 \underline{X}_t^{(m)}$  a vector of  $r$  high-frequency factors. Note that  $r = m - s$ , where  $s$  represents the rank reduction we are able to achieve within  $\underline{X}_t^{(m)}$ . In terms of parameter reduction, the unrestricted VAR in (2) requires  $p(m+1)^2$  coefficients to be estimated in the autoregressive matrices, whereas the VAR under reduced rank restrictions in (5) needs  $p(m+1) + r(m+1) + prm$  parameter estimates. As an example, assume  $p = 1$  and  $m = 20$ . Then, if  $r = 1, 2$  or  $3$ , there are, respectively, 62, 103 or 144 coefficients to be estimated in  $\underline{\Gamma}_1$  instead of 441 in  $\Gamma_1$ . While being reduced by a large amount, the number of parameters to be estimated remains considerable implying that the tests may still incur small sample size distortions. This is one of the reasons why we also consider bootstrap versions of these tests. Note that we do not require  $y_{t-1}$  to be included in the same transmission mechanism as the  $x$  variables.

**Remark 7** *There are several ways to justify the reduced rank feature of the autoregressive matrix  $B_{X(m)}$ . First, at the model representation level we may assume that, in the structural model (1),  $x$  follows an  $AR(q)$  process with  $q < m$  and that the last  $m - q$  elements of each  $X_{t-i}^{(m)}, i = 1, \dots, p$ , have a zero coefficient in the equation for  $y_t$ . Plugging these restrictions into (2) results in a reduced rank of  $B'_{X(m)}$  because the matrices  $\Gamma_i = A_c^{-1} A_i$  have the rank of  $A_i$ . Second, at the empirical level one can interpret the MF-VAR as an approximation of the VARMA obtained after block marginalization of a high-frequency VAR for each variable. In*

this situation, reduced rank matrices may empirically not be rejected by the data because of the combinations of many elements. Thus, a small number of dynamic factors can approximate more complicated (possibly nonlinear) dynamics. Finally, the way one typically restricts the MF-VAR in (2), i.e., assuming the high-frequency series to follow an ARX-process (Ghysels, 2016), actually implies a reduced rank representation. Looking slightly ahead, consider Equation (10) or the matrix in (12), which we will use in our Monte Carlo section, for the special case  $p = 1$ . It is obvious that for the first order MF-VAR,  $Z_t = \Gamma_1 Z_{t-1} + u_t$ , with

$$\Gamma_1 = \begin{pmatrix} \gamma_{1,1} & \gamma_{1,2} & \gamma_{1,3} & \dots & \gamma_{1,21} \\ \gamma_{2,1}^* & \rho^{20} & 0 & \dots & 0 \\ \gamma_{3,1}^* & \rho^{19} & \vdots & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{21,1}^* & \rho & 0 & \dots & 0 \end{pmatrix}$$

we have, due to the block of zeros, a reduced rank matrix  $B_{X(m)}$  with two (restricted coefficient) factors:

$$rk \left\{ \begin{pmatrix} \gamma_{1,2} & \gamma_{1,3} & \dots & \gamma_{1,21} \\ \rho^{20} & 0 & \dots & 0 \\ \rho^{19} & \vdots & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \rho & 0 & \dots & 0 \end{pmatrix} \right\} = rk \left\{ \begin{pmatrix} 1 & 0 \\ 0 & \rho^{20} \\ 0 & \rho^{19} \\ \vdots & \vdots \\ 0 & \rho \end{pmatrix} A A^{-1} \begin{pmatrix} \gamma_{1,2} & \gamma_{1,3} & \dots & \gamma_{1,21} \\ 1 & 0 & \dots & 0 \end{pmatrix} \right\} = 2,$$

where  $A$  is a  $2 \times 2$  full rank rotation matrix, the identity in this particular example. Hence, under  $H_A$ , e.g., two factors (although unrestricted in our setting) will capture the reduced rank feature. Note that we could impose (overidentifying) restrictions to retain the large block of zeros in  $\Gamma_1$ . In that sense, the model with  $r = 2$  would not be misspecified but “suboptimal”.

*On the other hand, with merely one factor we would not be able to impose the block of zeros implying the model to be misspecified. Under the null hypothesis the model is not misspecified with either one or two factors, because the test tries to capture some dynamics that is not there.*

Note that our focus is to compare the performance of different approaches including their bootstrap versions (whenever applicable) and that in this light we view reduced rank regressions as just one particular specification. In particular, next to the Bayesian MF-VAR and the *max*-test, our bootstrap implementation has also made the unrestricted MF-VAR (which is not subject to rank misspecification) a viable option to compare the reduced rank regressions with.

### 3.1.2 Testing for Granger Non-Causality

Given  $r$  and Assumption 6, under the condition that the high-frequency factors  $\delta' \underline{X}_t^{(m)}$  are observable, we can estimate (5) by OLS. Letting  $\hat{\underline{\Gamma}} = (\hat{\mu}, \hat{\gamma}_{\cdot,1}, \hat{\alpha})'$  denote the corresponding OLS estimates, we can test for Granger non-causality by defining  $R$  as the matrix that picks the set of coefficients we want to do inference on, i.e.,  $Rvec(\hat{\underline{\Gamma}})$ . For a general construction of the matrix  $R$  in the presence of several low- and high-frequency variables,<sup>6</sup> we refer the reader to Ghysels et al. (2016b). The Wald test is constructed as

$$\tilde{\xi}_W = \left[ Rvec(\hat{\underline{\Gamma}}) \right]' (R\hat{\Omega}R')^{-1} \left[ Rvec(\hat{\underline{\Gamma}}) \right], \quad (7)$$

with  $\hat{\Omega} = (W'W)^{-1} \otimes \hat{\Sigma}_\nu$ , where  $\hat{\Sigma}_\nu = \frac{1}{T} \sum_{t=p+1}^T \hat{\nu}_t \hat{\nu}_t'$ ,  $\hat{\nu}_t$  are the OLS residuals of (5), and  $W = (W_1, \dots, W_T)'$  is the regressor set consisting of  $W_t = (1, \underline{y}_t', \delta' \underline{X}_t^{(m)'})'$ .<sup>7</sup> As illustrated in Ghysels et al. (2016b),  $\xi_W$  is asymptotically  $\chi_{rank(R)}^2$ . A robust version of (7) is also implemented in the empirical section.

Of course, the factors are typically not observable and must be estimated (unless we im-

---

<sup>6</sup>Within the unrestricted MF-VAR in (2), though.

<sup>7</sup>A sample size correction, i.e., using  $\hat{\Sigma}_\nu = \frac{1}{T-K_W} \sum_{t=p+1}^T \hat{\nu}_t \hat{\nu}_t'$ , where  $K_W$  is the amount of elements in  $W$ , may alleviate size distortions in finite samples of tests using asymptotic critical values.

pose specific factors). Using estimated rather than observed factors in the regression (5) will obviously affect the distribution of  $\tilde{\xi}_W$ , certainly in small samples. For this reason we consider bootstrap versions of the tests. Before going into the details on the bootstrap we describe how we estimate or impose the factors.

### 3.1.3 Estimating the Factors

We use three ways to estimate the factors, canonical correlation analysis (CCA hereafter), partial least squares (PLS hereafter) and heterogeneous autoregressive (HAR hereafter) type restrictions. We briefly present the algorithms that are used to extract those factors.

CCA is based on analyzing the eigenvalues and corresponding eigenvectors of the matrix  $\hat{\Sigma}_{\tilde{X}^{(m)} \tilde{X}^{(m)}}^{-1} \hat{\Sigma}_{\tilde{X}^{(m)} \tilde{Z}} \hat{\Sigma}_{\tilde{Z} \tilde{Z}}^{-1} \hat{\Sigma}_{\tilde{Z} \tilde{X}^{(m)}}$  or, similarly,  $\hat{\Sigma}_{\tilde{X}^{(m)} \tilde{X}^{(m)}}^{-1/2} \hat{\Sigma}_{\tilde{X}^{(m)} \tilde{Z}} \hat{\Sigma}_{\tilde{Z} \tilde{Z}}^{-1} \hat{\Sigma}_{\tilde{Z} \tilde{X}^{(m)}} \hat{\Sigma}_{\tilde{X}^{(m)} \tilde{X}^{(m)}}^{-1/2}$ , a symmetric matrix. For a detailed discussion we refer the reader to Anderson (1951) or, for the application to common dynamics, to Vahid and Engle (1993). Note that  $\hat{\Sigma}_{ij}$  represents the empirical covariance matrix of processes  $i$  and  $j$ . Furthermore,  $\tilde{Z}$  and  $\tilde{X}^{(m)}$  indicate  $Z_t$  and  $\underline{X}_t^{(m)}$ , respectively, to be concentrated out by the variables that do not enter in the reduced rank regression, i.e., the intercept and  $y_{t-1}$ . Denoting by  $\hat{V} = (v_1, v_2, \dots, v_r)$ , with  $v_i' v_j = 1$  for  $i = j$  and 0 otherwise, the eigenvectors corresponding to the  $r$  largest eigenvalues of the matrix above, we obtain  $\hat{\alpha} = \hat{\Sigma}_{\tilde{Z} \tilde{Z}}^{-1} \hat{\Sigma}_{\tilde{Z} \tilde{X}^{(m)}} \hat{\Sigma}_{\tilde{X}^{(m)} \tilde{X}^{(m)}}^{-1/2} \hat{V}$  and  $\hat{\delta} = \hat{\Sigma}_{\tilde{X}^{(m)} \tilde{X}^{(m)}}^{-1/2} \hat{V}$ .

Note that the estimation of the eigenvectors obtained from the canonical correlation analysis in the matrices above may, however, perform poorly with high-dimensional systems, because inversions of the large variance matrices  $\hat{\Sigma}_{\tilde{Z} \tilde{Z}}^{-1}$  and  $\hat{\Sigma}_{\tilde{X}^{(m)} \tilde{X}^{(m)}}^{-1}$  are required. As an alternative to CCA we use a PLS algorithm similar to the one used in Cubadda and Hecq (2011) or Cubadda and Guardabascio (2012). In order to make the solution of this eigenvalue problem invariant to scale changes of individual elements, we compute the eigenvectors associated with the largest eigenvalues of the matrix  $\hat{D}_{\tilde{X}^{(m)} \tilde{X}^{(m)}}^{-1/2} \hat{\Sigma}_{\tilde{X}^{(m)} \tilde{Z}} \hat{D}_{\tilde{Z} \tilde{Z}}^{-1} \hat{\Sigma}_{\tilde{Z} \tilde{X}^{(m)}} \hat{D}_{\tilde{X}^{(m)} \tilde{X}^{(m)}}^{-1/2}$  with  $\hat{D}_{\tilde{X}^{(m)} \tilde{X}^{(m)}}$  and  $\hat{D}_{\tilde{Z} \tilde{Z}}$  being diagonal matrices having the diagonal elements of, respectively,  $\hat{\Sigma}_{\tilde{X}^{(m)} \tilde{X}^{(m)}}$  and  $\hat{\Sigma}_{\tilde{Z} \tilde{Z}}$  as



their entries. The computation of  $\hat{\alpha}$  and  $\hat{\delta}$  works in a similar fashion as with CCA-based factors.

Finally, we may *impose* the presence of  $r = 3p$  factors,<sup>8</sup> inspired by the HAR-model (Corsi, 2009). For  $i = 1, \dots, p$ :

$$\delta_i = \begin{pmatrix} \mathbf{0}_{3(i-1) \times m} \\ \hline 1 & \mathbf{0}_{1 \times (m-1)} \\ \hline \mathbf{1}_{1 \times (\frac{1}{4}m)} & \mathbf{0}_{(1 \times \frac{3}{4}m)} \\ & \mathbf{1}_{1 \times m} \\ & \mathbf{0}_{3(p-i) \times m} \end{pmatrix}' \Rightarrow \delta' \underline{X}_t^{(m)} = \begin{pmatrix} x_{t-1}^{(m)} \\ \sum_{i=0}^{\frac{1}{4}m-1} x_{t-1-i/m}^{(m)} \\ \sum_{i=0}^{m-1} x_{t-1-i/m}^{(m)} \\ \vdots \\ x_{t-p}^{(m)} \\ \sum_{i=0}^{\frac{1}{4}m-1} x_{t-p-i/m}^{(m)} \\ \sum_{i=0}^{m-1} x_{t-p-i/m}^{(m)} \end{pmatrix}. \quad (8)$$

For  $p = 1$  and  $m = 20$  this corresponds to

$$\delta' \underline{X}_t^{(m)} = \begin{pmatrix} x_{t-1}^{(20)} \\ \sum_{i=0}^4 x_{t-1-i/20}^{(20)} \\ \sum_{i=0}^{19} x_{t-1-i/20}^{(20)} \end{pmatrix} \equiv \begin{pmatrix} x_{t-1}^D \\ x_{t-1}^W \\ x_{t-1}^M \end{pmatrix}, \quad (9)$$

with  $x_t^D, x_t^W$  and  $x_t^M$  denoting daily, weekly and monthly measures, respectively.

**Remark 8** As noted in Ghysels and Valkanov (2012) and Ghysels et al. (2007), these HAR-type restrictions are a special case of MIDAS with step functions introduced in Forsberg and Ghysels (2007). Considering partial sums of regressors  $x$  as  $X_t(K, m) = \sum_{i=0}^K x_{t-i/m}^{(m)}$ , a MIDAS regression with  $M$  steps reads as  $y_t = \mu + \sum_{j=1}^M \beta_j X_t(K_j, m) + \epsilon_t$ , where  $K_1 < \dots < K_M$ . Alternatively, we could use MIDAS restrictions as introduced in Ghysels et al. (2004), for the difference between the latter and HAR-type restrictions has been shown to be small (Ghysels and Valkanov, 2012). However, step functions have the advantage of not requiring non-linear

---

<sup>8</sup>To ensure that  $r < m$  we assume that  $p < \frac{1}{3}m$  at this stage.

estimation methods, since the distributed lag pattern is approximated by a number of discrete steps, which simplifies the analysis. Furthermore, and more crucially in the context of this paper, implementing MIDAS restrictions and testing for Granger non-causality implies the well-known Davies (1987) problem, i.e., the parameters determining the MIDAS weights (see Ghysels et al., 2004 for details) are not identified under the null hypothesis.<sup>9</sup>

### 3.1.4 Bootstrap Tests

Despite the dimensionality reduction achieved by imposing the factor structure, a considerable number of parameters remains to be estimated for conducting the Wald tests such that the tests may still be subject to considerable small sample size distortions. Moreover, the estimation of the factors will have a major effect on the distribution of the Wald test statistic. We therefore also consider a bootstrap implementation of these tests to improve the properties of the tests in finite samples.

Let  $\hat{\underline{B}} = \hat{\Gamma}\hat{\delta}'$  denote the estimates of  $\underline{B}$  obtained by one of the reduced rank methods, and assume that  $\hat{\delta}$  is normalized such that its upper  $r \times r$  block is equal to the identity matrix.<sup>10</sup> The first bootstrap method we consider is a standard “unrestricted” bootstrap, that may have superior power properties in certain cases (cf. Paparoditis and Politis, 2005). Its algorithm looks as follows.

1. Obtain the residuals from the MF-VAR( $p$ ),  $\tilde{\nu}_t = Z_t - \hat{\mu} - \hat{\underline{B}}' \underline{Z}_t, t = p + 1, \dots, T$ .
2. Draw the bootstrap errors  $\nu_1^*, \dots, \nu_T^*$  with replacement from  $\tilde{\nu}_{y,p+1}, \dots, \tilde{\nu}_{y,T}$ .

---

<sup>9</sup>To properly test for Granger non-causality in this case, a grid for the weight specifying parameter vector has to be considered and the corresponding Wald tests for each candidate have to be computed. Subsequently, one can calculate the supremum of these tests (Davies, 1987) and obtain an ‘asymptotic  $p$ -value’ using bootstrap techniques (see Hansen, 1996 or Ghysels et al., 2007 for details). While this approach is feasible, it is computationally more demanding. Admittedly, HAR-type restrictions provide less flexibility than the MIDAS approach, yet their simplicity makes them very appealing from an applied perspective.

<sup>10</sup>This ensures that the rotation of the factors is the same for the original sample and the bootstrap sample, which in turn makes sure the correct bootstrap null hypothesis is imposed for the first bootstrap method. In the simulations we experimented with different normalizations which did not change the results.

3. Letting  $\underline{Z}_t^* = (Z_{t-1}^{*'}, \dots, Z_{t-p}^{*'})'$ , construct the bootstrap sample  $Z_t^*$  recursively as  $Z_t^* = \hat{\mu} + \hat{\underline{B}}' \underline{Z}_t^* + \nu_t^*, t = 1, \dots, T, \quad Z_0^* = 0$ . Note that the null hypothesis in the bootstrap is not imposed which has consequences for the next step.
4. Estimate the bootstrap equivalent of (5), where the factors are estimated using the same method as for the original sample, and obtain the bootstrap Wald test statistic  $\tilde{\xi}_W^*$ . As the “true” bootstrap parameters governing Granger causality are different from zero, we need to adapt the bootstrap Wald test such that it tests the correct null hypothesis. Observing that those parameters form a subset of the estimates in  $\underline{B}$  used in step 3, the appropriate Wald test statistic is  $\tilde{\xi}_W^* = \left[ \text{Rvec}(\hat{\underline{\Gamma}}^*) - \text{Rvec}(\hat{\underline{\Gamma}}) \right]' (\text{R}\hat{\Omega}^*\text{R}')^{-1} \left[ \text{Rvec}(\hat{\underline{\Gamma}}^*) - \text{Rvec}(\hat{\underline{\Gamma}}) \right]$ , where all quantities with a superscript “\*” are calculated analogously to their sample counterparts but then using the bootstrap sample.
5. Repeat steps 2 to 4  $B$  times, and calculate the bootstrap  $p$ -value as the proportion of bootstrap samples for which  $\tilde{\xi}_W^* > \tilde{\xi}_W$ .

While this bootstrap method has the advantage that it can be used for testing Granger causality in both directions (as only the matrix  $\text{R}$  needs to change), it also has some drawbacks. In particular, even with the dimensionality reduction provided by the reduced rank estimation, it still relies on the generation of a bootstrap sample based on an  $(m + 1)$ -dimensional MF-VAR( $p$ ) that requires  $p(m + 1) + r(m + 1) + pr m$  parameters to estimate. This may make the bootstrap unstable and prone to generate outlying samples, with potential size distortions or a loss of power as a result.

Therefore we consider a second bootstrap method that imposes the null hypothesis of no Granger causality and in doing so achieves a further significant reduction of dimensionality. This method is very similar to the parametric wild bootstrap proposed in Ghysels et al. (2016a). A downside of this method is that it requires separate bootstrap algorithms for testing in the two opposite directions. We describe the procedure here for the null hypothesis that  $X^{(m)}$  does not

Granger cause  $y$ , and comment on the test in the opposite direction below.

1. Estimate an  $\text{AR}(p)$  model for  $y$  and obtain the residuals,  $\tilde{u}_{y,t} = y_t - \hat{\mu} - \sum_{j=1}^p \hat{\rho}_{y,j} y_{t-j}$  for  $t = p+1, \dots, T$ .
2. Draw the bootstrap errors  $u_1^*, \dots, u_T^*$  with replacement from  $\tilde{u}_{y,p+1}, \dots, \tilde{u}_{y,T}$ .
3. Construct the bootstrap sample  $y_t^*$  recursively as  $y_t^* = \sum_{j=1}^p \hat{\rho}_{y,j} y_{t-1}^* + u_t^*$  for  $t = 1, \dots, T$  with  $y_0^* = 0$ , thus imposing the null of no Granger causality from  $X^{(m)}$  to  $y$ .
4. Letting  $X_t^{(m)*} = X_t^{(m)}$ , use the bootstrap sample to estimate the MF-VAR using the same estimation method as for the original estimator, and obtain the corresponding Wald test statistic  $\tilde{\xi}_W^*$ . As the null hypothesis of no Granger causality has been imposed in the bootstrap,  $\tilde{\xi}_W^*$  now has the same form as  $\tilde{\xi}_W$  in (7).

Note that only  $p$  parameters need to be estimated in order to generate the bootstrap sample, most likely making the procedure more stable than the unrestricted bootstrap above. A similar bootstrap procedure can be implemented for the reverse direction of causality by only resampling  $X_t^{(m)}$  while keeping  $y_t$  fixed. As one can resample  $X_t^{(m)}$  independently of  $y_t$ , it can be done in the original high-frequency as a univariate process. Said differently, it only requires fitting an  $\text{AR}(q)$  process to  $x_t^{(m)}$ , again achieving a large reduction of dimensionality.

Given the simplicity and apparent advantages of the second bootstrap method, one may wonder why one should consider the first, unrestricted, bootstrap method at all. However, in our simulation study we find that the unrestricted method has better size properties than the restricted method for testing causality from  $y$  to  $X^{(m)}$ , in particular in the presence of nowcasting causality.<sup>11</sup>

It is also possible in step 2 of both algorithms to use the wild bootstrap to be robust against heteroskedasticity. In this case the bootstrap errors are generated as  $u_t^* = \xi_t^* \tilde{u}_t$ , where we

---

<sup>11</sup>Simulation results explicitly comparing the two bootstrap methods are available upon request.

generate  $\xi_1^*, \dots, \xi_T^*$  as independent standard normal random variables. This would constitute the so-called “recursive wild bootstrap” scheme that Gonçalves and Kilian (2004) for univariate autoregressions and Brüggemann et al. (2016) for VAR models prove to be asymptotically valid under conditionally heteroskedastic errors. In particular, employing the wild bootstrap then allows us to replace the i.i.d. assumption in Assumption 2 with a martingale difference sequence assumption. We implement the wild bootstrap version in the empirical application in Section 5.

**Remark 9** *Kilian (1998) and Paparoditis (1996) prove the asymptotic validity of the unrestricted VAR bootstrap under Assumption 2. The validity of the restricted bootstrap can be derived from the results for AR(p) processes in Bose (1988), and the observation that conditioning is a valid approach in the absence of Granger causality. Van Giersbergen and Kiviet (1996) provide a detailed examination of the two different types of bootstraps in the context of autoregressive distributed lag models and demonstrate that the restricted bootstrap even works well in the absence of strong exogeneity. Dufour et al. (2006) apply the bootstrap to causality testing in VAR models and relax the assumption of i.i.d. error terms.<sup>12</sup>*

**Remark 10** *Gonçalves and Perron (2014) consider factor-augmented regression models, with factors estimated by principal components. In this setting the asymptotic impact of the factor estimation depends on which asymptotic framework is assumed. Under the asymptotic framework, in which factor estimation has a non-negligible effect on the limit distribution of the regression estimators, the bootstrap correctly mimics this effect and is therefore asymptotically valid. Consequently, the bootstrap provides a much more accurate approximation to the finite sample distribution of the regression estimators than the asymptotic approximation, that assumes a negligible effect of factor estimation. We expect the bootstrap to have similar properties in our setting where the factors are estimated using CCA or PLS.*

---

<sup>12</sup>If the test statistic (7) is asymptotically pivotal, we may expect the bootstrap to provide asymptotic refinements as well and, hence, reduce small sample size distortions (see Bose, 1988 or Horowitz, 2001).

## 3.2 Bayesian MF-VARs

### 3.2.1 Restricted MF-VARs

Ghysels (2016) discusses the issue of parsimony in MF-VAR models by specifying the high-frequency process in such a way as to allow a number of parameters that is independent from  $m$ . Indeed, while it seems reasonable to leave the equation for  $y_t$  unrestricted,<sup>13</sup> it is less clear for the remaining ones. Hence, let us make the following assumption (see Ghysels, 2016 or the numerical examples in Ghysels et al., 2016a).

**Assumption 11** *The high-frequency process  $x_t^{(m)}$  follows an  $AR(1)$  model with one lag of the low-frequency variable in the regressor set, i.e.,  $x_{t-i/m}^{(m)} = \mu_{i+2} + \rho x_{t-(i+1)/m}^{(m)} + \pi y_{t-1} + v_{(i+2)t}$  for  $i = 0, \dots, m-1$ .*

---

<sup>13</sup>In fact, viewed as single equation, it boils down to an unrestricted MIDAS model (Foroni et al., 2015) without contemporaneous observations of the high-frequency variable. MIDAS restrictions (Ghysels et al., 2004) can be imposed here as an alternative. However, doing so implies leaving the linear framework, which is needed to apply the auxiliary dummy variable approach presented below. Furthermore, even after having drawn the MIDAS hyperparameters, one still faces the aforementioned Davies (1987) problem when attempting to test for Granger non-causality from  $X^{(m)}$  to  $y$ .

Completing the system with the equation for  $y_t$  leads to the following restricted MF-VAR:

$$\begin{aligned}
\begin{pmatrix} y_t \\ x_t^{(m)} \\ x_{t-1/m}^{(m)} \\ \vdots \\ x_{t-(m-1)/m}^{(m)} \end{pmatrix} &= \begin{pmatrix} \mu_1 \\ \sum_{i=0}^{m-1} \rho^i \mu_{2+i} \\ \sum_{i=0}^{m-2} \rho^i \mu_{3+i} \\ \vdots \\ \mu_{m+1} \end{pmatrix} + \begin{pmatrix} \gamma_{1,1}^{(1)} & \gamma_{1,2}^{(1)} & \gamma_{1,3}^{(1)} & \cdots & \gamma_{1,m+1}^{(1)} \\ \pi \sum_{i=0}^{m-1} \rho^i & \rho^m & & & \\ \pi \sum_{i=0}^{m-2} \rho^i & \rho^{m-1} & & & \\ \vdots & \vdots & & \mathbf{0}_{m \times (m-1)} & \\ \pi & \rho & & & \end{pmatrix} \times \\
&\begin{pmatrix} y_{t-1} \\ x_{t-1}^{(m)} \\ x_{t-1-1/m}^{(m)} \\ \vdots \\ x_{t-1-(m-1)/m}^{(m)} \end{pmatrix} + \sum_{k=2}^p \begin{pmatrix} \gamma_{1,1}^{(k)} & \gamma_{1,2}^{(k)} & \cdots & \gamma_{1,m+1}^{(k)} \\ & \mathbf{0}_{m \times (m+1)} & & \end{pmatrix} \\
&\times \begin{pmatrix} y_{t-k} \\ x_{t-k}^{(m)} \\ \vdots \\ x_{t-k-(m-1)/m}^{(m)} \end{pmatrix} + \underbrace{\begin{pmatrix} v_{1t} \\ \sum_{i=0}^{m-1} \rho^i v_{(m+1-i)t} \\ \sum_{i=0}^{m-2} \rho^i v_{(m+1-i)t} \\ \vdots \\ v_{(m+1)t} \end{pmatrix}}_{v_t^*}, \tag{10}
\end{aligned}$$

where  $\gamma_{i,j}^{(k)}$  corresponds to the  $(i,j)$ -element of matrix  $\Gamma_k$  in (2). As for the error terms  $v_{it}$ ,  $i = 1, \dots, m+1$ , we set  $\mathbb{E}(v_{it}v_{it}) = \sigma_{HH}$ ,  $\mathbb{E}(v_{1t}v_{it}) = \sigma_{HL}$  for  $i = 2, \dots, m+1$ , and  $\mathbb{E}(v_{1t}v_{1t}) = \sigma_{LL}$ . Furthermore, each error term is assumed to possess a zero mean and to be normally distributed. Consequently,  $v_t^* \sim N(\mathbf{0}_{((m+1) \times 1)}, \Sigma_{v^*})$ , whereby we refer the reader to Ghysels (2016) for the exact composition of  $\Sigma_{v^*}$ .

### 3.2.2 The Auxiliary Dummy Variable Approach for MF Data

As pointed out by Carriero et al. (2011), Bayesian methods allow the imposition of restrictions such as the ones in Assumption 11, while also admitting influence of the data. Consequently, Bayesian shrinkage has become a standard tool when being faced with large-dimensional estimation problems such as large VARs (e.g., Bańbura et al., 2010, Kadiyala and Karlsson, 1997). As for MF-VARs, Ghysels (2016) describes a way to sample the MIDAS hyperparameters and subsequently formulates prior beliefs for the remaining parameters.<sup>14</sup> Once these hyperparameters are taken care of, the Bayesian analyses of mixed- and common-frequency VAR models are quite similar and, hence, traditional Bayesian VAR techniques (e.g., Kadiyala and Karlsson, 1997, Litterman, 1986) can be applied.

We follow the approach of Bańbura et al. (2010), which in turn is built on the work of Sims and Zha (1998), showing that adding a set of auxiliary dummy variables to the system is equivalent to imposing a normal inverted Wishart prior. The specification of prior beliefs is derived from the Minnesota prior in Litterman (1986), whereby we center the prior distributions of the coefficients in  $B$  around the restricted MF-VAR in (10):

$$\begin{aligned} \mathbb{E}[\gamma_{i,j}^{(k)}] &= \begin{cases} \rho^m & \text{if } i = j = k = 1 \\ \rho^{m+j-i} & \text{if } k = 1, j = 2, i > 1 \\ 0 & \text{else} \end{cases} , \\ \text{Var}[\gamma_{i,j}^{(k)}] &= \begin{cases} \phi \frac{\lambda^2}{k^2} \mathbb{S}_{LH} & \text{for } i = 1, j > 1 \\ \phi \frac{\lambda^2}{k^2} \mathbb{S}_{HL} & \text{for } j = 1, i > 1 \\ \frac{\lambda^2}{k^2} & \text{else} \end{cases} , \end{aligned} \tag{11}$$

where all  $\gamma_{i,j}^{(k)}$  are assumed to be a priori independent and normally distributed. The covariance

---

<sup>14</sup>McCracken et al. (2015) use a Sims-Zha shrinkage prior and the algorithm in Waggoner and Zha (2003) to solve the parameter proliferation problem with Bayesian estimation techniques. Bayesian methods within mixed-frequency VARs are also considered by Schorfheide and Song (2015). However, they use a latent high-frequency VAR instead of the mixed-frequency system à la Ghysels (2016).



matrix of the residuals is for now assumed diagonal and fixed, i.e.,  $\Sigma_u = \Sigma = \Sigma_d$  with  $\Sigma_d = \text{diag}(\sigma_L^2, \sigma_H^2, \dots, \sigma_H^2)$  of dimension  $m + 1$ . The tightness of the prior distributions around the AR(1) specification in Assumption 11 is determined by  $\lambda$ ,<sup>15</sup> the influence of low- on high-frequency data and vice versa is controlled by  $\phi$  and, finally,  $\mathbb{S}_{ij} = \frac{\sigma_i^2}{\sigma_j^2}, i, j = L, H$ , governs the difference in scaling between  $y$  and the  $x$ -variables. For  $\mu$  we take a diffuse prior.

**Remark 12** *As for the common-frequency case, the expressions for  $\text{Var}[\gamma_{i,j}^{(k)}]$  in (11) imply that more recent (low-frequency) lags provide more reliable information than more distant ones. However, due to the stacked nature of  $X_t^{(m)}$ , the coefficients could be shrunk according to their high-frequency time difference instead. This implies specifying the lag associated with each coefficient in fractions of the low-frequency time index:  $y_t$  and  $x_{t-1-i/m}^{(m)}$ , e.g., are  $1 + i/m$  low-frequency time periods apart such that the denominator in the corresponding coefficient's prior variance would equal  $(1 + i/m)^2$ .*

*However, in the context of Granger causality testing such a specification is problematic as the coefficients to test on are shrunk in "opposite directions":  $y_t$  and  $x_{t-1-1/2}^{(m)}$  are 1.5  $t$ -periods apart, whereas  $x_{t-1-1/2}^{(m)}$  and  $y_{t-1}$  are separated by only 0.5 low-frequency periods. Consequently, for a given  $\lambda$ , the coefficients are shrunk much more when testing for Granger causality from  $X^{(m)}$  to  $y$ , especially for large  $m$ . This makes it difficult to control the size of the respective tests in one or the other direction. The common-frequency handling of the prior variances in (11) mitigates this issue, although some loss in power has to be assumed.*

*Note that a treatment of the mixed-frequency nature of the variables along the lines outlined above may well be advantageous in other circumstances (e.g., forecasting) such that we present this approach in detail in Appendix A.<sup>16</sup>*

<sup>15</sup> $\lambda \approx 0$  results in the posterior coinciding with the prior, whereas  $\lambda = \infty$  causes the posterior mean to coincide with the OLS estimate of the unrestricted VAR in (2).

<sup>16</sup>When extending the approach we stick to the standard structure of the Minnesota prior in Litterman (1986). It is, however, also feasible to address the aforementioned issue by introducing, say, three hyperparameters:  $\lambda_1$  governing the prior tightness for the coefficients determining Granger causality from  $y$  to  $X_t^{(m)}$ ,  $\lambda_2$  for parameters controlling causality in the reverse direction and  $\lambda_3$  taking care of the remaining coefficients. This strategy is,

Given the prior beliefs specified before, the analysis is very similar to the one in Bańbura et al. (2010). In short, let us write the MF-VAR as  $Z = \underline{Z}B^* + E$ , where  $\underline{Z} = (\underline{Z}_1^\mu, \dots, \underline{Z}_T^\mu)'$  with  $\underline{Z}_t^\mu = (\underline{Z}_t', 1)'$ ,  $Z = (Z_1, \dots, Z_T)'$ ,  $E = (u_1, \dots, u_T)'$  and  $B^* = (B', \mu)'$ . Then, one can show that augmenting the model by two dummy variables,  $Y_d$  and  $X_d$ , is equivalent to imposing a normal inverted Wishart prior that satisfies the moments in (11). Finally, estimating the augmented model by ordinary least squares gives us the posterior mean of the coefficients, on which we can do inference as outlined in the next subsection.

Note that  $X_d$  is, in fact, identical to the matrix for the common-frequency VAR (Bańbura et al., 2010);  $Y_d$ , however, is slightly different due to the prior means being centered around the restricted VAR in (10):

$$\underbrace{Y_d}_{((m+1)(p+1)+1) \times (m+1)} = \begin{pmatrix} \rho^m \sigma_L / \lambda & 0 & 0 & \dots & 0 \\ 0 & \rho^m \sigma_H / \lambda & \rho^{m-1} \sigma_H / \lambda & \dots & \rho \sigma_H / \lambda \\ \mathbf{0}_{((m+1)p-2) \times (m+1)} \\ \text{diag}(\sigma_L, \sigma_H, \dots, \sigma_H) \\ \mathbf{0}_{1 \times (m+1)} \end{pmatrix},$$

Augmenting the model is then achieved by setting  $Z_* = (Z', Y_d')'$  and  $\underline{Z}_* = (\underline{Z}', X_d')'$ .

### 3.2.3 Testing for Granger Non-Causality

In terms of analyzing the Granger non-causality testing behavior within the Bayesian MF-VAR, the auxiliary dummy variable approach is quite appealing, as it provides a closed-form solution for the posterior mean of the coefficients. It is thus straightforward to compare the testing behavior with the ones from alternative approaches using the Wald test.<sup>17</sup> The latter

---

however, a rather straightforward extension of the theory presented in Appendix A such that we do not present it here.

<sup>17</sup>There is a large sample correspondence between classical Wald and Bayesian posterior odds tests (Andrews, 1994). For certain choices of the prior distribution, the posterior odds ratio is approximately equal to the Wald statistic. Andrews (1994) shows that for any significance level  $\alpha$  there exist priors such that the aforementioned

is derived analogously to the one in (7), and is, given Assumption 11, also asymptotically  $\chi^2_{rank(R)}$ -distributed.

**Remark 13** *Because the parameter vector  $vec(B^*)$  is not assumed fixed but random, construction of the test statistic, and especially its interpretation, need to be treated with special care. We consider Bayesian confidence intervals, in particular highest posterior density (HPD) confidence intervals (Bauwens et al., 2000), the tightness of which is expressed by  $\alpha$ , not coincidentally the same letter that denotes the significance level in the frequentist’s framework. Here, it is interpreted such that  $(1 - \alpha)\%$  of the probability mass falls within the respective interval. In other words, the probability that a model parameter falls within the bounds of the interval is equal to  $(1 - \alpha)\%$ . The interval centered at the modal value for unimodal symmetric posterior densities is then called the HPD (see, e.g., Bauwens et al., 2000). The connection to Wald tests is now immediate using the equivalence between confidence intervals and respective test statistics, whereby similar care is demanded when interpreting results.*

### 3.3 Benchmark Models

#### 3.3.1 Low-Frequency VAR

Before the introduction of MIDAS regression models, high-frequency variables were usually aggregated to the low frequency in order to obtain a common frequency for all variables appearing in a regression (Silvestrini and Veredas, 2008 or Marcellino, 1999). Likewise for systems, a monthly variable, for example, was usually aggregated to, say, the quarterly frequency such that a VAR could be estimated in the resulting common low frequency. Formally,  $x_t = W(L^{1/m})x_t^{(m)}$ , where  $W(L^{1/m})$  denotes a high-frequency lag polynomial of order  $A$ , i.e.,  $W(L^{1/m})x_t^{(m)} = \sum_{i=0}^A w_i x_{t-i/m}^{(m)}$  (Silvestrini and Veredas, 2008).<sup>18</sup> As far as testing for Granger

---

correspondence holds, and vice versa.

<sup>18</sup>This generic specification nests the two dominating aggregation schemes in the literature, Point-in-Time ( $A = 0, w_0 = 1$ ) and Average sampling ( $A = m - 1, w_i = 1/m \forall i$ ), where the former is usually applied to stock

non-causality is concerned, we can rely on the asymptotic Wald statistic in (7), where the set of regressors, the matrices  $\hat{\Omega}$  and  $R$  as well as the coefficient matrix are suitably adjusted.<sup>19</sup>

**Remark 14** *Naturally, such temporal aggregation leads to a great reduction in parameters. After all, each set of  $m$  high-frequency variables per  $t$ -period is aggregated into one low-frequency observation. Of course, this decrease in parameters comes at the cost of disregarding information embedded in the high-frequency process. As argued in Miller (2011), if the aggregation scheme employed is different from the true one underlying the DGP, potentially crucial high-frequency information will be forfeited. Additionally, aggregating a high-frequency variable may lead to “spurious” (non-)causality in the common low-frequency setup (Breitung and Swanson, 2002), since causality is a property which is not invariant to temporal aggregation (Marcellino, 1999 or Sims, 1971).*

### 3.3.2 The *max*-test

Independently and simultaneously to this work, Ghysels et al. (2016a) have developed a new Granger non-causality testing framework, whose parsimonious structure makes it very appealing in a situation, where  $m$  is large relative to the sample size. In short, the idea is to focus on the first line of the MF-VAR in (2), but rather than estimating that (U)-MIDAS equation (Forni et al., 2015) and test for Granger non-causality in the direction from  $X^{(m)}$  to  $y$ , the authors propose to compute the OLS estimates of  $\beta_j$  in the following  $h$  separate regression models:

$$y_t = \mu + \sum_{k=1}^q \alpha_{k,j} y_{t-k} + \beta_{j+1} x_{t-1-j/m}^{(m)} + v_j, \quad j = 0, \dots, h-1,$$

---

and the latter to flow variables. In view of the high-frequency variable we consider in our empirical application, the natural logarithm of bipower variation, we focus on Average sampling in this paper.

<sup>19</sup>We only consider the asymptotic Wald test in our simulations as it turns out to be correctly sized; hence there is no need for the bootstrap.

whereby  $h$  needs to be set “sufficiently large” (to achieve  $h > pm$ ). The corresponding *max*-test statistic is then the properly scaled and weighted maximum of  $\{\hat{\beta}_1^2, \dots, \hat{\beta}_h^2\}$ . Although it has a non-standard limit distribution under  $H_0$ , an asymptotic  $p$ -value may be obtained in a similar way as when overcoming the Davies (1987) problem (see Remark 8). Testing for Granger non-causality in the reverse direction works analogously in the following regression model:

$$y_t = \mu + \sum_{k=1}^q \alpha_{k,j} y_{t-k} + \sum_{k=1}^h \beta_{k,j} x_{t-1-k/m}^{(m)} + \gamma_j x_{t+j/m}^{(m)} + v_j, \quad j = 1, \dots, l.$$

Again, the corresponding *max*-test statistic is the maximum of  $\{\hat{\gamma}_1^2, \dots, \hat{\gamma}_l^2\}$  scaled and weighted properly.<sup>20</sup> Ghysels et al. (2016a) show that, in a mixed-frequency VAR à la Ghysels (2016), the sequential procedure underlying the *max*-test leads to correct identification of the test in the presence or absence of Granger causality. Hence, the *max*-test is consistent against causality.

### 3.3.3 Unrestricted VARs

Finally, we can attempt to estimate the full MF-VAR in (2) ignoring the possibility that the amount of parameters may be too large to perform accurate estimations or adequate Granger non-causality tests. To this end we just estimate the MF-VAR using ordinary least squares. In this sense the comparison is related to the one of U-MIDAS (Foroni et al., 2015) and MIDAS regression models for large  $m$ . We can test for Granger non-causality using the Wald statistic in (7), adequately adjusting  $W$ ,  $\hat{\Omega}$ ,  $R$  and  $B$ . We also consider a bootstrap version of the unrestricted MF-VAR, which we expect to alleviate size distortions, but which cannot solve power issues due to the proliferation of parameters. Indeed, as shown by Davidson and MacKinnon (2006), the power of a bootstrap test can only equal the size-corrected power of the original test; as the parameter proliferation causes the original Wald test to be grossly oversized, the

---

<sup>20</sup>A MIDAS polynomial (Ghysels et al., 2004) on  $\sum_{k=1}^h \beta_{k,j} x_{t-1-k/m}^{(m)}$  may be imposed to increase parsimony. Note that the aforementioned Davies problem does not occur due to testing for Granger non-causality by looking at the OLS estimates of  $\gamma_j$ .

amount of size-correcting needed implies the remaining power to be low.

## 4 Monte Carlo Simulations

In order to assess the finite sample performance of our different parameter reduction techniques, we conduct a Monte Carlo experiment. In light of our empirical investigation we set  $m = 20$ , i.e., as in a month/ working day-example.<sup>21</sup> Furthermore, we start by investigating the case where  $p = 1$  and keep the analysis of higher lag orders for future research.

As far as investigating the size of our Granger non-causality tests is concerned, we assume that the data are generated as a mixed-frequency white noise process, i.e.,  $Z_t = u_t$ .

**Remark 15** *Additionally, we have considered three alternative DGPs for size, all based on the restricted VAR in (10): (i)  $\gamma_{1,i} = \gamma_{i,1}^* = 0$  ('diagonal'), (ii)  $\gamma_{1,i} = 0$  and  $\gamma_{i,1}^* \neq 0$  ('only Granger non-causality from  $X^{(m)}$  to  $y$ ') and (iii)  $\gamma_{i,1}^* = 0$  and  $\gamma_{1,i} \neq 0$  ('only Granger non-causality from  $y$  to  $X^{(m)}$ ')  $\forall i = 2, \dots, 21$ , where the parameters  $\gamma_{1,i}$  and  $\gamma_{i,1}^*$  refer to (12). However, with the respect to the respective testing direction, the outcomes do not differ qualitatively and quantitative differences are very small. Results are available upon request.*

To analyze power we generate two different DGPs that are closely related to the restricted VAR in (10), i.e.,  $Z_t = \Gamma_P Z_{t-1} + u_t$  with

$$\Gamma_P = \begin{pmatrix} \gamma_{1,1} & \gamma_{1,2} & \gamma_{1,3} & \cdots & \gamma_{1,21} \\ \gamma_{2,1}^* & \rho^{20} & 0 & \cdots & 0 \\ \gamma_{3,1}^* & \rho^{19} & \vdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{21,1}^* & \rho & 0 & \cdots & 0 \end{pmatrix}, \quad (12)$$

---

<sup>21</sup>See Section 5 for a justification of the time-invariance of  $m$  within this setup. The merits of our methods should hold for larger values of  $m$ ; for smaller values, however, we may assume our approaches to perform worse. Indeed, the smaller  $m$  becomes, the less is the need to reduce the amount of parameters via (possibly misspecified) reduction techniques.

where  $\gamma_{1,1} = 0.5$ ,  $\rho = 0.6$  and  $\gamma_{1,j} = \beta w_{j-1}^*(-0.01)$  for  $j > 1$ . Note that  $w_i(\psi) = \frac{\exp(\psi i^2)}{\sum_{i=1}^{20} \exp(\psi i^2)}$  corresponds to the two-dimensional exponential Almon lag polynomial with the first parameter set to zero (Ghysels et al., 2007). Now, for the first power DGP we simply set  $w_i^*(\psi) = w_i(\psi)$ , whereas in the second power DGP  $w_i^*(\psi) = w_i(\psi) - \overline{w(\psi)}$  with the horizontal bar symbolizing the arithmetic mean. As far as  $\gamma_{j,1}^*$  is concerned we take  $\gamma_{j,1}^* = \gamma_{j,1}$  (first power DGP) and  $\gamma_{j,1}^* = \gamma_{j,1} - \overline{\gamma_{\cdot,1}}$  (second power DGP) for  $j = 2, \dots, 21$ , whereby  $\gamma_{j,1} = (\gamma_{j-1,1})^{1.11}$  for  $j = 3, \dots, 21$ .<sup>22</sup> As an example, Figure 1 plots  $\beta w_{j-1}^*$  and  $\gamma_{j,1}^*$  for  $\beta = 2$  and  $\gamma_{2,1} = 0.25$ .

[INSERT FIGURE 1 HERE]

**Remark 16** *Due to the zero-mean feature of  $w_{j-1}^*(\psi)$  and  $\gamma_{j,1}^*, j = 2, \dots, 21$ , in the second power DGP, we expect the presence of Granger causality to be “hidden” when Average sampling the high-frequency variable (see Section 3.3.1). It is thus a situation, in which classical temporal aggregation is expected not to preserve the causality patterns in the data (Marcellino, 1999). The first power DGP serves as a benchmark in the sense that we do not a priori expect one approach to be better or worse than the others.*

For the methods in Section 3 we analyze size and power for  $T = 50, 250$  and  $500$ , corresponding to roughly 4, 21 and 42 years of monthly data. Note that an additional 100 monthly observations are used to initialize the process. As far as the error term is concerned, we assume  $u_t \stackrel{i.i.d.}{\sim} N(\mathbf{0}_{(21 \times 1)}, \Sigma_u)$ , where  $\Sigma_u$  has the same structure as the covariance matrix of the restricted VAR in (10) with  $\sigma_{LL} = 0.5$  and  $\sigma_{HH} = 1$ .<sup>23</sup>

For CCA and PLS we consider  $r = 1, 2$  and leave the analysis of higher factor dimensions for further research. Recall from Remark 7 that under  $H_0$  the model is not misspecified with

<sup>22</sup>The parameter values have been chosen to mimic part of the structure of the restricted VAR in (10) and to ensure stability of the system. In the first DGP late  $x_{t-1}$ -observations have a larger impact on  $y_t$ , and positive values of  $y_{t-1}$  increase  $x$  towards the end of the current period, but hardly have an impact at the beginning. In the second power DGP a zero-mean feature is imposed on the coefficients  $w_{j-1}^*(\psi)$  and  $\gamma_{j,1}^*$  without changing the general pattern of their evolution.

<sup>23</sup>These values resemble the data in the empirical section, where  $\hat{\sigma}_{LL} = 0.55$  and  $\hat{\sigma}_{HH} = 1.09$ .

either amount of factors. Under  $H_A$ , however, the model is misspecified for  $r = 1$ , but not for  $r = 2$ . In that sense, we implicitly analyze the impact of misspecification, stemming from choosing  $r$  too small, on the behavior of the respective test statistics. As far as the Bayesian approach is concerned, we experimented with various values for the hyperparameter and found  $\lambda = 0.175$  to be a good choice. Furthermore, we approximate  $\rho$  by running a corresponding univariate regression. For the *max*-test we take  $q = 1$  and  $h = l = m = 20$ .

We consider different versions of our size and power DGPs by varying the values of  $\sigma_{HL}$  and  $\beta$ . To be more precise, we choose  $\sigma_{HL} = 0, -0.05, -0.1$  when analyzing size properties of our tests in both directions.<sup>24</sup> A value of zero implies the absence of nowcasting causality, whereas the other two values imply some degree of correlation between the low- and high-frequency series. For both power DGPs we fix  $\sigma_{HL} = -0.05$  though. When investigating Granger causality testing from  $X^{(m)}$  to  $y$  in the first power DGP we consider  $\beta = 0.5, 0.8, 2$ ; in the second power DGP we only take  $\beta = 0.5$ . For causality in the reverse direction we take  $\beta = 0.5$  and  $\gamma_{2,1} = 0.2$  in both power DGPs, because the results do not seem sensitive in this respect.

The figures in the tables below represent the percentage amount of rejections at  $\alpha = 5\%$ .<sup>25</sup> All figures are based on 2,500 replications and are computed using GAUSS12. For the bootstrap versions we use  $B = 499$ .

## 4.1 Size

Tables 1 and 2 contain the size results for Granger non-causality tests within the following approaches: the low-frequency VAR (LF), the unrestricted VAR (UNR), reduced rank restrictions using canonical correlations analysis (CCA) or partial least squares (PLS), and with three

---

<sup>24</sup> $\hat{\sigma}_{HL}$  turns out to be equal to  $-0.22$  in our empirical analysis, yet we decrease its value here so as to achieve positive-definiteness of  $\Sigma_u$ . We checked the robustness of our outcomes to  $\sigma_{HL}$  being positive (we chose the corresponding values  $0.05$  and  $0.1$ ) as this situation is more common in economic applications. However, the results remain broadly unchanged and are available upon request.

<sup>25</sup>Outcomes for  $\alpha = 10\%$  and  $\alpha = 1\%$  are available upon request.



imposed factors using the HAR model (HAR), the *max*-test (*max*) and the Bayesian mixed-frequency VAR (BMF). With respect to CCA and PLS, the outcomes for  $r = 2$  turn out to be qualitatively very similar to the  $r = 1$  case, showing that the methods are apparently not affected by the presence of misspecification. Consequently, we only show the outcomes for  $r = 1$  to save space. Furthermore, aside from the results of the standard, i.e., asymptotic, Wald tests, we only display the outcomes for the best performing bootstrap variant, which are denoted with a superscript ‘\*’ (e.g., CCA\* is the bootstrap CCA-based Wald test). In the case of testing causality in the direction from the high- to the low-frequency series this turns out to be the bootstrap that imposes the null hypothesis, whereas the unrestricted variant dominates for the reverse direction.<sup>26</sup> For both methods we set the lag length within the bootstrap equal to  $p = 1$ .

When testing causality from  $y$  to  $X^{(m)}$  size distortions may occur due to computing a joint test on  $mp$  parameters from  $m$  different equations in the system. In order to address this issue, we take the maximum of the Wald statistics computed equation by equation. We denote these tests by adding a subscript ‘ $b$ ’, e.g., CCA <sub>$b$</sub> . For the tests with asymptotic critical values we apply a Bonferroni correction to control the size under multiple testing (Dunn, 1961). In similar spirit as in White (2000), the bootstrap implementations of these tests automatically provide an implicit Bonferroni-type correction for multiple testing, as we calculate the corresponding maximum of the single-equation Wald tests within each bootstrap iteration and compare that to the maximum of the original tests. We will refer to all these tests as ‘Bonferroni-type’ tests to avoid confusion with the *max*-test, and present the corresponding outcomes in Table 3.

[INSERT TABLE 1 HERE]

Let us start by analyzing the testing behavior from  $X^{(m)}$  to  $y$ . First, LF has size close to the nominal one, which is not surprising given that the flat aggregation scheme is correct in this

---

<sup>26</sup>The unrestricted bootstrap is seriously oversized when testing causality from  $X^{(m)}$  to  $y$ , while for the reverse direction the bootstrap under the null is mildly oversized for small samples. Hence, we recommend the use of the restricted bootstrap for testing causality from  $X^{(m)}$  to  $y$ , but using the unrestricted bootstrap for the opposite direction.

white noise DGP (no matter the value of  $\sigma_{HL}$ ). The unrestricted VAR, however, incurs some size distortions for small  $T$  due to parameter proliferation. Reduced rank restrictions based on CCA and PLS yield considerable size distortions, whereby the imposed HAR-type factor structure delivers good results (despite being slightly oversized for  $T = 50$ ). Finally, *max* and BMF show almost perfect size, whereby the former is marginally under- and the latter marginally oversized.

Turning to the outcomes of the bootstrap versions we observe that, independent of the value of  $\sigma_{HL}$ , even for  $T = 50$  the actual size of the Wald tests within the unrestricted VAR and reduced rank restriction models with CCA- or HAR-type factors is very close to the nominal one, such that the bootstrap tests clearly dominate their asymptotic counterparts. For PLS-based factors this conclusion only holds in the absence of nowcasting causality, and size distortions arise quickly as we increase the contemporaneous correlation between  $X^{(m)}$  and  $y$ .

[INSERT TABLE 2 HERE]

[INSERT TABLE 3 HERE]

Many of the aforementioned statements carry over to testing causality in the reverse direction: LF delivering nearly perfect size (as it is the correct model under the null), *max* performing very well (being only marginally oversized for small  $T$ ) and BMF being a bit oversized (by a slightly larger degree than in Table 1).<sup>27</sup> The size distortions incurred by UNR, CCA, PLS and HAR are, however, amplified. The use of the Bonferroni-corrected  $\text{UNR}_b$ ,  $\text{CCA}_b$ ,  $\text{PLS}_b$  and  $\text{HAR}_b$  can only mitigate this effect, yet not fully eradicate it. The bootstrap version does eliminate the size distortions, though: actual size of  $\text{UNR}^*$ ,  $\text{PLS}^*$  and  $\text{HAR}^*$  becomes oftentimes very close to 5%; only  $\text{CCA}^*$  remains a bit oversized for  $T = 50$ . The Bonferroni-type tests  $\text{UNR}_b^*$ ,  $\text{CCA}_b^*$ ,  $\text{PLS}_b^*$  and  $\text{HAR}_b^*$  all have size very close to the nominal one for all sample sizes.

---

<sup>27</sup>Strangely, BMF seems to be rejected less and less for growing  $\sigma_{HL}$ , with the effect being strongest for small  $T$ .

To sum up, the asymptotic Granger non-causality tests with size close to the nominal one are LF, *max* and BMF to some degree. As far as the bootstrap tests are concerned, UNR\*, CCA\*, PLS\* (with at most "mild" nowcasting causality) and HAR\* when testing the direction from  $X^{(m)}$  to  $y$ , and UNR\*, CCA\*, PLS\* and HAR\* as well as their Bonferroni-type counterparts when testing the reverse direction deliver good size outcomes. Consequently, we focus on these cases when analyzing power (and when dealing with real data in Section 5).

## 4.2 Power

Tables 4 and 5 contain the corresponding outcomes under the alternative hypothesis. Recall that for the direction from  $X^{(m)}$  to  $y$  we consider three different values of  $\beta$  within the first power DGP,  $\beta = 0.5, 0.8$  and  $2$ . For the second power DGP we fix  $\beta = 0.5$ . For the reverse direction we always keep  $\gamma_{2,1} = 0.2$ .

[INSERT TABLE 4 HERE]

Let us again start with the direction from the high- to the low-frequency variable and first focus on the outcomes for the first power DGP, i.e., the top three blocks of Table 4. Observe (i) how power reaches one asymptotically for all approaches, and (ii) how larger values of  $\beta$  imply an increase in the rejection frequencies for a given  $T$ . Naturally, for a large enough  $\beta$  all tests have power equal to one, irrespective of the sample size (as nearly happens for  $\beta = 2$ ). So, in order to compare our different tests let us focus on  $\beta = 0.5$ .

The asymptotic Wald test within the low-frequency VAR still performs very well, with the highest rejection frequency for  $T = 50$ . Note, however, that the Granger causality feature in the data does not get averaged out by temporal aggregation in the first power DGP. The bootstrapped version of the Wald test after HAR-type restrictions have been imposed in a reduced rank regression perform almost as well as LF. *max*, BMF and PLS\* appear to fall a bit short, but catch up quickly as either  $T$  or  $\beta$  grows. A similar observation holds for UNR\*

and  $CCA^*$ , whereby they seem to be markedly less powerful for small  $T$ .

These conclusions generally carry over to the second power DGP, with two exceptions: first and foremost, the outcomes of LF show that Average sampling of  $X^{(m)}$  annihilates the causality feature between the series for all sample sizes (power actually almost coincides with the nominal size of 5%). Second, the rejection frequencies are lower than they were in the corresponding setting of the first power DGP.

[INSERT TABLE 5 HERE]

Focusing on the direction from  $y$  to  $X^{(m)}$ , it suffices to analyze the results of the first power DGP; the aforementioned statements regarding the second power DGP carry over to the reverse direction of causality being tested. It turns out that both, *max* and BMF, have higher power than the bootstrapped Wald tests corresponding to the reduced rank restrictions approaches. Recall, though, that BMF and *max* were somewhat oversized, especially for small  $T$ , which may inflate their power. Also, so far we did not consider the Bonferroni-type tests. Given the way in which *max* is constructed it seems more natural to compare it to the Bonferroni-type counterparts of the bootstrap tests. Indeed, both approaches rely on a statistic that is computed as the maximum of a set of test statistics.<sup>28</sup> We find that for a given  $T$  almost all approaches are more powerful than *max*. The fact that  $UNR_b^*$ ,  $PLS_b^*$  and  $HAR_b^*$  are all slightly undersized for small  $T$ , while *max* is marginally oversized, even strengthens these conclusions. Also note that the bootstrap tests based on the reduced rank restrictions seem to be slightly more powerful than  $UNR_b^*$ . Hence, while the bootstrap is able to correct size of the unrestricted MF-VAR approach, the proliferation of parameters continues to have a negative effect on power.

Combining the power and size outcomes, it seems that BMF,  $HAR^*$ , *max* (for large enough  $\beta$  when  $T$  is small) and  $PLS^*$  (in the absence of nowcasting causality) are the dominant Granger

---

<sup>28</sup>There is a difference in terms of the underlying regression to obtain the various test statistics, of course. Ghysels et al. (2016a) consider univariate MIDAS regressions with leads, whereas the Bonferroni-type tests are based on estimated coefficients from different equations, i.e., the ones for  $X^{(m)}$ , of the corresponding system.

non-causality testing approaches as far as the direction from  $X^{(m)}$  to  $y$  is concerned. For the reverse direction of causality  $CCA_b^*$ ,  $PLS_b^*$  and  $HAR_b^*$  as well as  $UNR_b^*$  and  $max$  (though both to a lesser degree) appear superior. Despite the competitiveness of our proposed tests, asymptotic or bootstrapped, in the situation at hand, one should keep in mind that the underlying parameter reduction techniques are ultimately ad hoc. Hence, in contrast to the  $max$ -test and testing within the unrestricted VAR, which do not involve possibly misspecified dimension reduction approaches, our tests can ultimately not have an asymptotic power of one. As such, the good performance of our proposed tests may be specific to our simulation setup and should not be generalized to all possible causality patterns for which our imposed restrictions may not hold.

## 5 Application

We apply the approaches described in Section 3 to a MF-VAR consisting of the monthly growth rate of the U.S. industrial production index (*ipi* hereafter), a measure of business cycle fluctuations, and the logarithm of daily bipower variation (*bv* hereafter) of the S&P500 stock index, a realized volatility measure more robust to jumps than the realized variance. While the degree to which macroeconomic variables can help to predict volatility movements has been investigated widely in the literature (see Schwert, 1989b, Hamilton and Gang, 1996, or Engle and Rangel, 2008, among others), the reverse, i.e., whether the future path of the economy can be predicted using return volatility, has been granted comparably little attention (examples are Schwert, 1989a, Mele, 2007, and Andreou et al., 2000). Instead of using an aggregate measure of volatility such as, e.g., monthly realized volatilities (Chauvet et al., 2015) or monthly GARCH estimated variances, we use daily bipower variation computed on 5-minute returns obtained from the database in Heber et al. (2009). With the *bv*-series being available at a higher frequency than most indicators of business cycle fluctuations, we obtain the mixed-frequency framework analyzed in this paper.

The sample covers the period from January 2000 to June 2012 yielding a sample size of  $T = 150$ . We take  $m = 20$  as it is the maximum amount of working days that is available in *every* month throughout the sample we deal with.<sup>29</sup> Consequently, we have  $BV_t^{(20)} = (bv_t^{(20)}, bv_{t-1/20}^{(20)}, \dots, bv_{t-19/20}^{(20)})'$ . Figure 2 plots the data.

[INSERT FIGURE 2 HERE]

Table 6 contains the outcomes of Granger non-causality tests for all approaches discussed in Section 3. As mentioned before, we disregard the cases that showed considerable size distortions in the Monte Carlo analysis. Note that a lag length of  $p = 1$  and 2 is considered (the same for the bootstrap) and that the numbers represent  $p$ -values (in percentages). For reduced rank restrictions using CCA and PLS we consider one up to three factors, whereby we only show the outcomes for  $r = 1$  and  $r = 2$  for representational ease;  $r = 3$  gives similar results. For the non-bootstrap tests, except the *max*-test and the Wald-test within the Bayesian mixed-frequency VAR,<sup>30</sup> we also consider a heteroscedasticity consistent variant of (7) by computing a robust estimator of  $\Omega$  (see Ravikumar et al., 2000) to account for the potential presence of a time varying multivariate process:  $\hat{\Omega}_R = T((W'W)^{-1} \otimes I_{m+1})\hat{S}_0((W'W)^{-1} \otimes I_{m+1})$ , where  $\hat{S}_0 = \frac{1}{T} \sum_{t=1}^T (W_t \otimes \hat{u}_t)'(W_t \otimes \hat{u}_t)$ .

For the bootstrap tests we achieve the robustness to heteroskedasticity by implementing the wild bootstrap version as described in Section 3.1.4.

[INSERT TABLE 6 HERE]

---

<sup>29</sup>Whenever a month contains more than 20 working days we disregard the corresponding amount of days at the beginning of the month. In June 2012, e.g., there are 21 working days such that we do not consider June 1. For May 2012 we disregard the first three days. An alternative (balanced) strategy would have been to take the maximum number of days in a *particular* month (i.e., 23, usually in July, August or October) and to create additional values for non-existing days in other months whenever necessary. As far as the treatment of daily data is concerned we have also taken  $bv_t^{(20)} = bv_{t-1/20}^{(20)}$  when there are no quotations for  $bv_t^{(20)}$ .

<sup>30</sup>Note that, as discussed in Section 3.2.2, there is a one-to-one correspondence between the OLS estimator of the augmented model and the prior setting we consider in our Bayesian framework. As we did not discuss a prior specification that corresponds to the robust version of the Wald test we disregard from that test variant here.

If we consider the test statistics that perform best in our Monte Carlo experiment, the outcomes above clearly point towards Granger-causality from  $BV^{(20)}$  to  $ipi$ . This result supports Andreou et al. (2000) concluding that "[...] volatilities may also be useful [...] indicators for both the growth and volatility of industrial production" (p. 15). The situation is less clear cut when looking at Granger-causality from the low-frequency variable to the high-frequency volatility measures. Indeed, *max* and most Bonferroni-type tests seem to reject the null of no Granger-causality, whereas the joint tests do not. However, the higher power obtained on the maximum of the individual Wald-type tests would favor the presence of Granger-causality in the direction from business cycle movements to financial uncertainties as well. A more careful investigation, out of the scope of this paper, could be done on different subsamples in order to analyze whether, e.g., periods before, during or after the financial crises lead to similar conclusions.

## 6 Conclusion

We investigate Granger non-causality testing in a mixed-frequency VAR, where the mismatch between the sampling frequencies of the variables under consideration is large, causing estimation and inference to be potentially problematic. To avoid this issue we discuss two parameter reduction techniques in detail, reduced rank restrictions and a Bayesian MF-VAR approach, and compare them to (i) a common low-frequency VAR, (ii) the *max*-test approach and (iii) the unrestricted VAR in terms of their Granger non-causality testing behavior. To further improve their finite sample test properties we also consider two bootstrap variants for the reduced rank regression approaches (and the unrestricted VAR).

For both directions of causality we find a different set of tests that result in the best Granger non-causality testing behavior. For the direction from the high- to the low-frequency series, standard testing within the Bayesian mixed-frequency VAR, the *max*-test of Ghysels et al.

(2016a), and the restricted bootstrap version of the Wald test in a reduced rank regression after HAR-type factors have been imposed or PLS-type factors have been computed, the latter of which being restricted to nowcasting non-causality (Götz and Hecq, 2014), perform best. For the reverse direction, the unrestricted bootstrap variants of the Bonferroni-corrected Wald tests within the following models dominate: the unrestricted VAR and reduced rank regressions with CCA-, PLS- or HAR-based factors.

An application investigating the presence of a causal link between business cycle fluctuations and uncertainty in financial markets illustrates the practical usefulness of these approaches. While Granger causality from uncertainty in financial markets to business cycle fluctuations was clearly supported by the data, evidence for causality in the reverse direction only comes from a subset of the tests, yet the more powerful ones according to our simulation results.

## Acknowledgements

We want to particularly thank Eric Ghysels for many fruitful discussions, comments and suggestions on earlier versions of the paper. Moreover, we thank two anonymous referees, one of them deserving special gratitude for providing us with MATLAB code for the *max*-test. We also thank Daniela Osterrieder, Lenard Lieb, Jörg Breitung, Roman Liesenfeld, Jan-Oliver Menz, Klemens Hauzenberger, Martin Mandler and participants of the Workshop on Advances in Quantitative Economics in Maastricht 2013, the CEIS Seminar in Rome 2013, the 33<sup>rd</sup> ISF in Seoul 2013, the 24<sup>th</sup> (EC)<sup>2</sup> conference in Cyprus 2013, the 22<sup>th</sup> SNDE in New York 2014, the 8<sup>th</sup> ECB Workshop on Forecasting Techniques in Frankfurt 2014, the Graduate School Research Seminar in Cologne 2014, the 68<sup>th</sup> ESEM in Toulouse 2014, the 8<sup>th</sup> CFE in Pisa 2014 and the Norges Bank Research Seminar in Oslo 2015 for useful suggestions and comments on earlier versions of the paper. The third author would like to thank the Netherlands Organisation for Scientific Research (NWO) for financial support.



## References

- Anderson, T. W., 1951, Estimating linear restrictions on regression coefficients for multivariate normal distributions. *The Annals of Mathematical Statistics* 22, 327–351.
- Andreou, E., D. R. Osborn and M. Sensier, 2000, A comparison of the statistical properties of financial variables in the USA, UK and Germany over the business cycle. *Manchester School* 68, 396–418.
- Andrews, D. W. K., 1994, The large sample correspondence between classical hypothesis tests and Bayesian posterior odds tests. *Econometrica* 62, 1207–1232.
- Bañbura, M., D. Giannone and L. Reichlin, 2010, Large Bayesian vector autoregressions. *Journal of Applied Econometrics* 25, 71–92.
- Bauwens, L., M. Lubrano and J.-F. Richard, 2000, Bayesian inference in dynamic econometric models. Oxford University Press.
- Bose, A., 1988, Edgeworth correction by bootstrap in autoregressions. *Annals of Statistics* 16, 1709–1722.
- Breitung, J. and N. R. Swanson, 2002, Temporal aggregation and spurious instantaneous causality in multiple time series models. *Journal of Time Series Analysis* 23, 651–666.
- Brüggemann, R., C. Jentsch and C. Trenkler, 2016, Inference in VARs with conditional heteroskedasticity of unknown form. *Journal of Econometrics* 191, 69–85.
- Carriero, A., G. Kapetanios and M. Marcellino, 2011, Forecasting large datasets with bayesian reduced rank multivariate models. *Journal of Applied Econometrics* 26, 735–761.
- Cavaliere, G., A. Rahbek and A. M. R. Taylor, 2012, Bootstrap determination of the co-integration rank in vector autoregressive models. *Econometrica* 80, 1721–1740.

- Chauvet, M., Z. Senyuz and E. Yoldas, 2015, What does financial volatility tell us about macroeconomic fluctuations? *Journal of Economic Dynamics and Control* 52, 340–360.
- Clements, M. and A. B. Galvão, 2008, Macroeconomic forecasting with mixed-frequency data. *Journal of Business & Economic Statistics* 26, 546–554.
- Clements, M. and A. B. Galvão, 2009, Forecasting U.S. output growth using leading indicators: an appraisal using MIDAS models. *Journal of Applied Econometrics* 24, 1187–1206.
- Corsi, F., 2009, A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* 7, 174–196.
- Cubadda, G. and B. Guardabascio, 2012, A medium-N approach to macroeconomic forecasting. *Economic Modelling* 29, 1099–1105.
- Cubadda, G. and A. Hecq, 2011, Testing for common autocorrelation in data rich environments. *Journal of Forecasting* 30, 325–335.
- Davidson, R. and J. G. MacKinnon, 2006, The power of bootstrap and asymptotic tests. *Journal of Econometrics* 133, 421–441.
- Davies, R. B., 1987, Hypothesis testing when a nuisance parameter is present only under the alternatives. *Biometrika* 74, 33–43.
- Dufour, J.-M., D. Pelletier and E. Renault, 2006, Short run and long run causality in time series: inference. *Journal of Econometrics* 132, 337–362.
- Dufour, J.-M. and E. Renault, 1998, Short run and long run causality in time series: theory. *Econometrica* 66, 1099–1126.
- Dunn, O. J., 1961, Multiple comparisons among means. *Journal of the American Statistical Association* 56, 52–64.

- Engle, R. F. and J. G. Rangel, 2008, The spline-garch model for low-frequency volatility and its global macroeconomic causes. *Review of Financial Studies* 21, 1187–1222.
- Eraker, B., C. W. J. Chiu, A. T. Foerster, T. B. Kim and H. D. Seoane, 2015, Bayesian mixed frequency VARs. *Journal of Financial Econometrics* 13, 698–721.
- Foroni, C., E. Ghysels and M. Marcellino, 2013, Mixed-frequency vector autoregressive models, in: T. B. Fomby, L. Kilian and A. Murphy, (Eds.), *Advances in Econometrics*, Vol. 32. Emerald Group Publishing Limited, pp. 247–272.
- Foroni, C. and M. Marcellino, 2014, Mixed frequency structural models: identification, estimation, and policy analysis. *Journal of Applied Econometrics* 29, 1118–1144.
- Foroni, C., M. Marcellino and C. Schumacher, 2015, Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials. *Journal of the Royal Statistical Society Series A* 178, 57–82.
- Forsberg, L. and E. Ghysels, 2007, Why do absolute returns predict volatility so well? *Journal of Financial Econometrics* 5, 31–67.
- Ghysels, E., 2016, Macroeconomics and the reality of mixed frequency data. *Journal of Econometrics*, forthcoming.
- Ghysels, E., J. B. Hill and K. Motegi, 2016a, Simple Granger causality tests for mixed frequency data. Working paper, Dept. of Economics, University of North Carolina at Chapel Hill.
- Ghysels, E., J. B. Hill and K. Motegi, 2016b, Testing for Granger causality with mixed frequency data. *Journal of Econometrics* 192, 207–230.
- Ghysels, E. and J. I. Miller, 2015, Testing for cointegration with temporally aggregated and mixed-frequency time series. *Journal of Time Series Analysis* 36, 797–816.

- Ghysels, E., P. Santa-Clara and R. Valkanov, 2004, The midas touch: Mixed data sampling regression models. CIRANO Working Papers 2004s-20, CIRANO.
- Ghysels, E., P. Santa-Clara and R. Valkanov, 2006, Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics* 131, 59–95.
- Ghysels, E., A. Sinko and R. Valkanov, 2007, Midas regressions: further results and new directions. *Econometric Reviews* 26, 53–90.
- Ghysels, E. and R. Valkanov, 2012, Forecasting volatility with MIDAS, in: L. Bauwens, C. Hafner and S. Laurent, (Eds.), *Handbook of Volatility Models and Their Applications*, Chapter 16. John Wiley & Sons, Inc., Hoboken, NJ, USA, pp. 383–401.
- Gonçalves, S. and L. Kilian, 2004, Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. *Journal of Econometrics* 123, 89–120.
- Gonçalves, S. and B. Perron, 2014, Bootstrapping factor-augmented regression models. *Journal of Econometrics* 182, 156–173.
- Götz, T. B. and A. Hecq, 2014, Nowcasting causality in mixed frequency vector autoregressive models. *Economics Letters* 122, 74–78.
- Götz, T. B., A. Hecq and J.-P. Urbain, 2013, Testing for common cycles in non-stationary VARs with varied frequency data, in: T. B. Fomby, L. Kilian and A. Murphy, (Eds.), *Advances in Econometrics*, Vol. 32. Emerald Group Publishing Limited, pp. 361–393.
- Götz, T. B., A. Hecq and J.-P. Urbain, 2014, Forecasting mixed frequency time series with ECM-MIDAS models. *Journal of Forecasting* 33, 198–213.
- Granger, C. W. J., 1969, Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37, 424–438.

- Granger, C. W. J., 1988, Some recent development in a concept of causality. *Journal of Econometrics* 39, 199–211.
- Granger, C. W. J. and J.-L. Lin, 1995, Causality in the long run. *Econometric Theory* 11, 530–536.
- Hamilton, J. D. and L. Gang, 1996, Stock market volatility and the business cycle. *Journal of Applied Econometrics* 11, 573–93.
- Hansen, B. E., 1996, Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64, 413–30.
- Heber, G., A. Lunde, N. Shephard and K. Sheppard, 2009, Oxford-man institute’s realized library (version 0.2). Oxford-Man Institute, University of Oxford.
- Horowitz, J. L., 2001, The bootstrap, in: J. J. Heckman and E. E. Leamer, (Eds.), *Handbook of Econometrics*, Vol. 5, Chapter 52. North Holland Publishing, Amsterdam, pp.3159–3228.
- Kadiyala, K. R. and S. Karlsson, 1997, Numerical methods for estimation and inference in Bayesian VAR-models. *Journal of Applied Econometrics* 12, 99–132.
- Kilian, L., 1998, Small-sample confidence intervals for impulse response functions. *Review of Economics and Statistics* 80, 218–230.
- Kuzin, V., M. Marcellino, and C. Schumacher, 2011, MIDAS vs. mixed-frequency VAR: now-casting GDP in the euro area. *International Journal of Forecasting* 27, 529–542.
- Litterman, R. B., 1986, Forecasting with Bayesian vector autoregressions - Five years of experience. *Journal of Business & Economic Statistics* 4, 25–38.
- Lütkepohl, H., 1993, Testing for causation between two variables in higher dimensional VAR models, in: H. Schneeweiss, Hans und K. F. Zimmermann, (Eds.), *Studies in Applied Econometrics*. Springer-Verlag, Heidelberg, pp. 75–91.

- Marcellino, M., 1999, Some consequences of temporal aggregation in empirical analysis. *Journal of Business & Economic Statistics* 17, 129–136.
- Marcellino, M. and C. Schumacher, 2010, Factor MIDAS for Nowcasting and Forecasting with Ragged-Edge Data: a Model Comparison for German GDP. *Oxford Bulletin of Economics and Statistics* 72, 518–550.
- McCracken, M. W., M. T. Owyang and T. Sekhposyan, 2015, Real-Time forecasting with a large, mixed frequency, Bayesian VAR. Working Paper 2015-30, Federal Reserve Bank of St. Louis.
- Mele, A., 2007, Asymmetric stock market volatility and the cyclical behavior of expected returns. *Journal of Financial Economics* 86, 446–478.
- Miller, J. I., 2011, Conditionally efficient estimation of long-run relationships using mixed-frequency time series. Working Paper 1103, Department of Economics, University of Missouri.
- Miller, J. I., 2014, Mixed-frequency cointegrating regressions with parsimonious distributed lag structures. *Journal of Financial Econometrics* 12, 584–614.
- Paparoditis, E., 1996, Bootstrapping autoregressive and moving average parameter estimates of infinite order vector autoregressive processes. *Journal of Multivariate Analysis* 57, 277–296.
- Paparoditis, E. and D. N. Politis, 2005, Bootstrap hypothesis testing in regression models. *Statistics & Probability Letters* 74, 356–365.
- Park, T. and G. Casella, 2008, The Bayesian Lasso. *Journal of American Statistical Association* 103, 681–686.
- Ravikumar, B., S. Ray and N. E. Savin, 2000, Robust Wald tests in SUR systems with adding-up restrictions. *Econometrica* 68, 715–720.

- Schorfheide, F. and D. Song, 2015, Real-time forecasting with a mixed-frequency VAR. *Journal of Business & Economics Statistics* 33, 366–380.
- Schwert, G. W., 1989a, Business cycles, financial crises, and stock volatility. *Carnegie-Rochester Conference Series on Public Policy* 31, 83–125.
- Schwert, G. W., 1989b, Why does stock market volatility change over time? *Journal of Finance* 44, 1115–53.
- Silvestrini, A. and D. Veredas, 2008, Temporal aggregation of univariate and multivariate time series models: a survey. *Journal of Economic Surveys* 22, 458–497.
- Sims, C. A., 1971, Discrete approximations to continuous time distributed lags in econometrics. *Econometrica* 39, 545–563.
- Sims, C. A. and T. Zha, 1998, Bayesian methods for dynamic multivariate models. *International Economic Review* 39, 949–68.
- Vahid, F. and Engle, R. F., 1993, Common trends and common cycles. *Journal of Applied Econometrics* 8, 341–360.
- Van Giersbergen, N. P. A. and J. F. Kiviet, 1996, Bootstrapping a stable AD model: weak vs strong exogeneity. *Oxford Bulletin of Economics and Statistics* 58, 631–656.
- Vogel, C. R., 2002, Computational methods for inverse problems. *Society for Industrial and Applied Mathematics*, Philadelphia.
- Waggoner, D. F. and T. Zha, 2003, A Gibbs sampler for structural vector autoregressions. *Journal of Economic Dynamics and Control* 28, 349–366.
- White, H., 2000, A reality check for data snooping. *Econometrica* 68, 1097–1126.

## 7 Tables



Table 1: Size of Granger Non-Causality Tests from  $X^{(m)}$  to  $y$

$X^{(m)}$ to $y$	$T$	LF	UNR	CCA	PLS	HAR	$max$	BMF	UNR*	CCA*	PLS*	HAR*
$\sigma_{HL} = 0$	50	5.4	13.4	18.8	11.4	6.9	4.3	5.7	5.0	4.6	4.9	5.2
	250	4.7	6.4	28.6	12.4	5.4	4.4	5.2	4.9	4.7	5.3	5.3
	500	4.9	5.6	29.2	12.8	5.2	4.1	5.3	5.1	5.2	6.2	5.3
$\sigma_{HL} = -0.05$	50	5.3	12.6	20.9	16.7	6.6	4.6	5.6	4.7	5.6	8.2	5.0
	250	4.8	5.8	28.3	17.1	5.5	4.4	5.2	4.7	4.8	9.2	5.4
	500	4.8	6.2	30.0	16.9	4.4	4.2	5.3	5.6	5.2	8.5	4.5
$\sigma_{HL} = -0.1$	50	5.0	12.5	20.0	32.6	6.3	4.6	5.0	4.8	4.8	19.6	4.8
	250	4.6	7.2	28.8	35.7	5.5	4.8	5.1	5.8	5.8	23.2	5.2
	500	4.7	5.4	28.9	36.5	5.3	4.5	5.2	4.8	6.3	23.0	5.2

Note: For testing Granger non-causality from  $X^{(m)}$  to  $y$  the figures represent percentages of rejections of the asymptotic Wald test statistics and the restricted bootstrap variants (indicated with a superscript \*) of the following approaches: The low-frequency VAR (LF), the unrestricted approach (UNR), reduced rank restrictions with one factor using canonical correlations analysis (CCA) or partial least squares (PLS), and with three imposed factors using the HAR model (HAR), the  $max$ -test ( $max$ ) and the Bayesian mixed-frequency approach (BMF). The lag length of the estimated VARs is equal to one. The underlying DGP is  $Z_t = u_t$ , where the variance-covariance matrix of the error term is equal to  $\Sigma_u$  with  $\sigma_{HL} = 0, -0.05, -0.1$ .

Table 2: Size of Granger Non-Causality Tests from  $y$  to  $X^{(m)}$

$y$ to $X^{(m)}$	$T$	LF	UNR	CCA	PLS	HAR	$max$	BMF	UNR*	CCA*	PLS*	HAR*
$\sigma_{HL} = 0$	50	5.7	91.1	82.0	58.6	59.8	6.1	7.9	4.6	8.8	5.0	5.2
	250	5.5	11.5	11.8	10.6	10.2	4.6	10.4	5.2	5.3	5.0	5.3
	500	5.2	6.7	7.4	6.8	6.6	5.9	7.8	4.6	4.8	4.6	4.8
$\sigma_{HL} = -0.05$	50	5.6	92.4	83.8	62.0	60.6	5.7	4.8	5.6	9.4	4.9	5.6
	250	5.4	10.9	13.2	11.9	10.3	4.8	8.9	4.7	4.5	4.4	4.8
	500	5.0	7.4	9.2	8.7	7.3	5.5	7.3	5.4	5.4	5.3	5.2
$\sigma_{HL} = -0.1$	50	5.7	92.1	87.1	64.2	58.5	5.7	0.5	5.5	9.2	4.0	4.4
	250	4.7	11.0	20.5	14.8	10.5	5.0	5.4	5.2	6.3	5.1	5.3
	500	5.3	7.4	13.8	10.7	7.2	5.5	5.0	5.3	5.7	4.2	5.2

Note: For testing Granger non-causality from  $y$  to  $X^{(m)}$  the figures represent percentages of rejections of the asymptotic Wald test statistics and the unrestricted bootstrap variants (indicated with a superscript \*) of the following approaches: The low-frequency VAR (LF), the unrestricted approach (UNR), reduced rank restrictions with one factor using canonical correlations analysis (CCA) or partial least squares (PLS), and with three imposed factors using the HAR model (HAR), the  $max$ -test ( $max$ ) and the Bayesian mixed-frequency approach (BMF). The lag length of the estimated VARs is equal to one. The underlying DGP is  $Z_t = u_t$ , where the variance-covariance matrix of the error term is equal to  $\Sigma_u$  with  $\sigma_{HL} = 0, -0.05, -0.1$ .

Table 3: Size of Granger Non-Causality Tests from  $y$  to  $X^{(m)}$  (Bonferroni)

$y$ to $X^{(m)}$	$T$	UNR $_b$	CCA $_b$	PLS $_b$	HAR $_b$	BMF $_b$	UNR $_b^*$	CCA $_b^*$	PLS $_b^*$	HAR $_b^*$
$\sigma_{HL} = 0$	50	16.9	17.3	16.4	13.9	2.9	3.4	4.6	3.9	3.7
	250	12.7	13.6	13.1	12.7	5.1	4.6	5.0	5.0	4.9
	500	12.4	12.4	12.1	12.2	4.4	5.0	5.0	4.6	4.9
$\sigma_{HL} = -0.05$	50	17.9	18.9	18.3	15.2	2.1	3.8	4.8	4.0	4.0
	250	13.4	14.5	14.4	12.8	4.8	5.4	5.2	5.4	5.4
	500	12.3	13.7	14.8	12.5	4.0	5.0	5.2	5.0	4.6
$\sigma_{HL} = -0.1$	50	17.2	20.2	21.6	15.4	0.4	3.3	5.0	4.3	4.4
	250	12.0	15.4	17.0	12.2	3.3	4.9	5.3	4.8	4.7
	500	11.9	14.9	17.0	11.7	3.3	5.4	5.2	5.1	5.2

Note: For testing Granger non-causality from  $y$  to  $X^{(m)}$  the figures represent percentages of rejections of the Bonferroni-type tests (indicated with a subscript 'b') for both the asymptotic Wald test statistics and the unrestricted bootstrap variants (indicated with a superscript '\*') of the following approaches: The unrestricted approach (UNR), reduced rank restrictions with one factor using canonical correlations analysis (CCA) or partial least squares (PLS), and with three imposed factors using the HAR model (HAR) and the Bayesian mixed-frequency approach (BMF). For the low-frequency VAR (LF) and the *max*-test (*max*) the Bonferroni-type test is not applicable. The lag length of the estimated VARs is equal to one. The underlying DGP is  $Z_t = u_t$ , where the variance-covariance matrix of the error term is equal to  $\Sigma_u$  with  $\sigma_{HL} = 0, -0.05, -0.1$ .

Table 4: Power of Granger Non-Causality Tests from  $X^{(m)}$  to  $y$

$X^{(m)}$ to $y$	$T$	LF	$max$	BMF	UNR*	CCA*	PLS*	HAR*
Power-DGP 1, $\beta = 0.5$	50	64.1	39.6	52.8	23.1	12.7	54.0	62.6
	250	99.9	100.0	99.8	99.8	96.0	98.7	100.0
	500	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Power-DGP 1, $\beta = 0.8$	50	93.6	79.4	92.2	54.6	26.6	77.9	93.0
	250	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	500	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Power-DGP 1, $\beta = 2$	50	100.0	100.0	100.0	100.0	85.3	98.4	100.0
	250	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	500	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Power-DGP 2, $\beta = 0.5$	50	5.8	14.5	21.9	9.7	8.0	25.1	22.3
	250	5.0	75.9	79.2	71.2	58.8	79.7	89.4
	500	4.7	98.6	99.0	98.3	87.6	97.3	99.8

Note: For testing Granger non-causality from  $X^{(m)}$  to  $y$  the figures represent percentages of rejections of the asymptotic Wald test statistics of the low-frequency VAR (LF), the  $max$ -test ( $max$ ) and the Bayesian mixed-frequency approach (BMF). Furthermore, it contains percentages of rejections of the restricted bootstrap variants (indicated with a superscript \*\*) of the unrestricted approach (UNR), reduced rank restrictions with one factor using canonical correlations analysis (CCA) or partial least squares (PLS), and with three imposed factors using the HAR model (HAR). The lag length of the estimated VARs is equal to one. The underlying DGP is  $Z_t = \Gamma_P Z_{t-1} + u_t$  with  $\Gamma_P$  defined in (12), where the variance-covariance matrix of the error term is equal to  $\Sigma_u$  with  $\sigma_{HL} = -0.05$ . In power-DGP 1 the Granger causality determining coefficients do not sum up to zero, whereas they do so in power-DGP 2. See Section 4 for details.  $\beta = 0.5, 0.8, 2$  in power-DGP 1 and  $\beta = 0.5$  in power-DGP 2.

Table 5: Power of Granger Non-Causality Tests from  $y$  to  $X^{(m)}$

$y$ to $X^{(m)}$	$T$	LF	$max$	BMF	UNR*	CCA*	PLS*	HAR*	UNR*	CCA*	PLS*	HAR*
Power-DGP 1	50	6.9	8.0	15.8	5.6	9.9	6.2	6.2	6.8	11.4	10.7	10.8
	250	19.9	24.2	31.0	17.7	20.0	20.3	19.3	51.0	54.3	54.7	53.8
	500	36.4	53.2	49.1	39.3	44.7	43.9	42.4	83.1	85.2	85.2	84.6
Power-DGP 2	50	4.7	7.9	18.4	5.7	10.6	5.7	5.3	5.9	10.0	8.9	8.0
	250	5.0	22.9	25.1	13.5	14.6	14.3	14.3	32.2	34.7	34.2	32.6
	500	5.2	49.9	33.7	26.6	29.9	28.4	27.7	57.3	62.2	62.2	59.3

Note: For testing Granger non-causality from  $y$  to  $X^{(m)}$  the figures represent percentages of rejections of the asymptotic Wald test statistics of the low-frequency VAR (LF), the  $max$ -test ( $max$ ) and the Bayesian mixed-frequency approach (BMF). Furthermore, it contains percentages of rejections of the unrestricted bootstrap variants (indicated with a superscript '\*') of the unrestricted approach (UNR), reduced rank restrictions with one factor using canonical correlations analysis (CCA) or partial least squares (PLS), and with three imposed factors using the HAR model (HAR). For the latter set of approaches, it also shows the outcomes for the Bonferroni-type tests (indicated with a subscript 'b'). The lag length of the estimated VARs is equal to one. The underlying DGP is  $Z_t = \Gamma_P Z_{t-1} + u_t$  with  $\Gamma_P$  defined in (12), where the variance-covariance matrix of the error term is equal to  $\Sigma_u$  with  $\sigma_{HL} = -0.05$ . In power-DGP 1 the Granger causality determining coefficients do not sum up to zero, whereas they do so in power-DGP 2. See Section 4 for details.  $\beta = 0.5$  and  $\gamma_{2,1} = 0.2$  in both power-DGPs.

Table 6: Testing for Granger Causality between  $BV^{(20)}$  and  $ipi$

$BV^{(20)}$ to $ipi$	$p = 1$		$p = 2$	
	Wald	Robust	Wald	Robust
LF	0.0	0.2	0.1	2.1
<i>max</i>	0.1	./.	0.0	./.
BMF	14.4	./.	21.8	./.
UNR*	15.0	17.4	7.3	8.7
CCA1*	2.9	7.9	12.4	20.4
CCA2*	14.5	16.2	19.2	25.1
PLS1*	0.1	0.7	0.2	1.9
PLS2*	0.6	2.8	0.1	1.4
HAR*	1.2	2.2	1.5	2.4
$ipi$ to $BV^{(20)}$	$p = 1$		$p = 2$	
	Wald	Robust	Wald	Robust
LF	0.1	1.5	0.4	5.0
<i>max</i>	4.4	./.	1.1	./.
BMF	23.5	./.	42.5	./.
UNR*	28.9	34.2	25.8	26.4
CCA1*	25.8	33.5	20.5	38.4
CCA2*	35.9	46.1	17.6	34.3
PLS1*	15.8	24.6	11.4	24.4
PLS2*	18.3	26.9	10.3	18.0
HAR*	19.2	27.0	11.4	18.6
UNR <sub>b</sub> *	5.4	8.9	20.0	22.9
CCA1 <sub>b</sub> *	4.1	10.5	7.8	16.9
CCA2 <sub>b</sub> *	5.0	10.8	7.7	15.1
PLS1 <sub>b</sub> *	1.5	4.3	1.5	9.5
PLS2 <sub>b</sub> *	1.2	5.2	1.1	4.0
HAR <sub>b</sub> *	2.3	6.5	6.2	14.1

Note: For testing Granger non-causality between  $BV^{(20)}$  and  $ipi$  the figures represent  $p$ -values (in percentages) of the asymptotic Wald test statistics, the restricted bootstrap variants for the direction from  $BV^{(20)}$  to  $ipi$  and the unrestricted bootstrap variants for the direction from  $ipi$  to  $BV^{(20)}$  (bootstrap variants are indicated with a superscript '\*', the Bonferroni-type tests with a subscript 'b'). For all approaches but the *max*-test (*max*) and the Bayesian MF-VAR (BMF) a robustified version is computed as well. The lag length of the estimated VARs is equal to one or two. For CCA and PLS the results for one and two factors are displayed.

## 8 Figures

Figure 1: Parameter values for  $2w_{j-1}^*(-0.01)$  and  $\gamma_{j,1}^*$  ( $\gamma_{2,1} = 0.25$ )

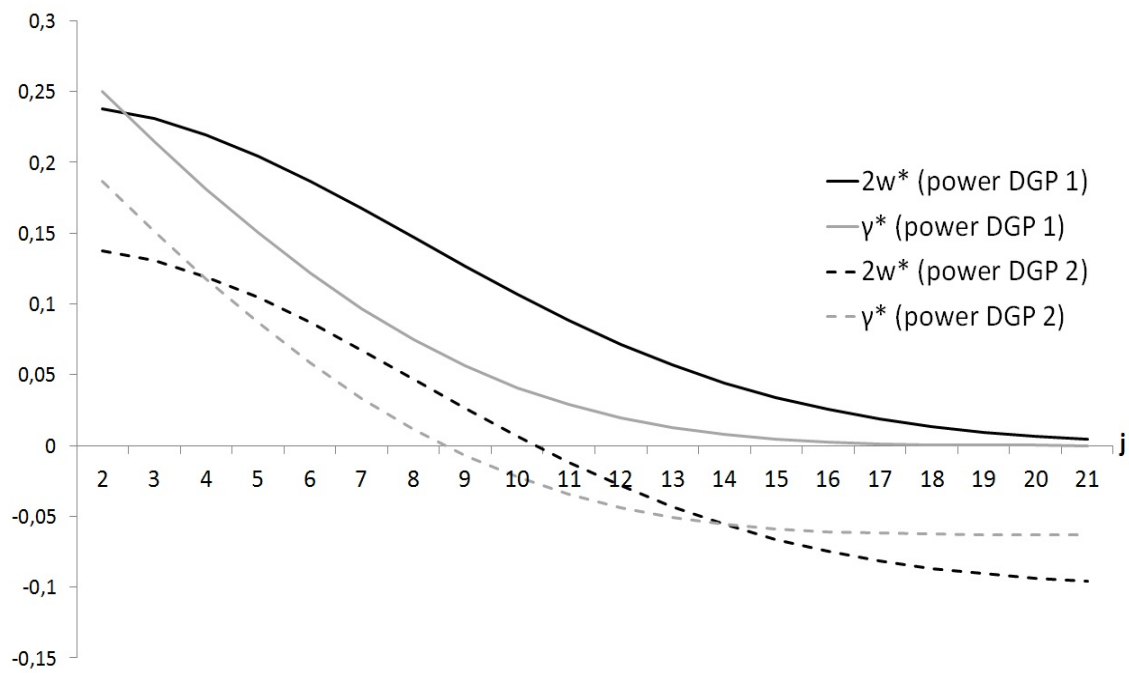
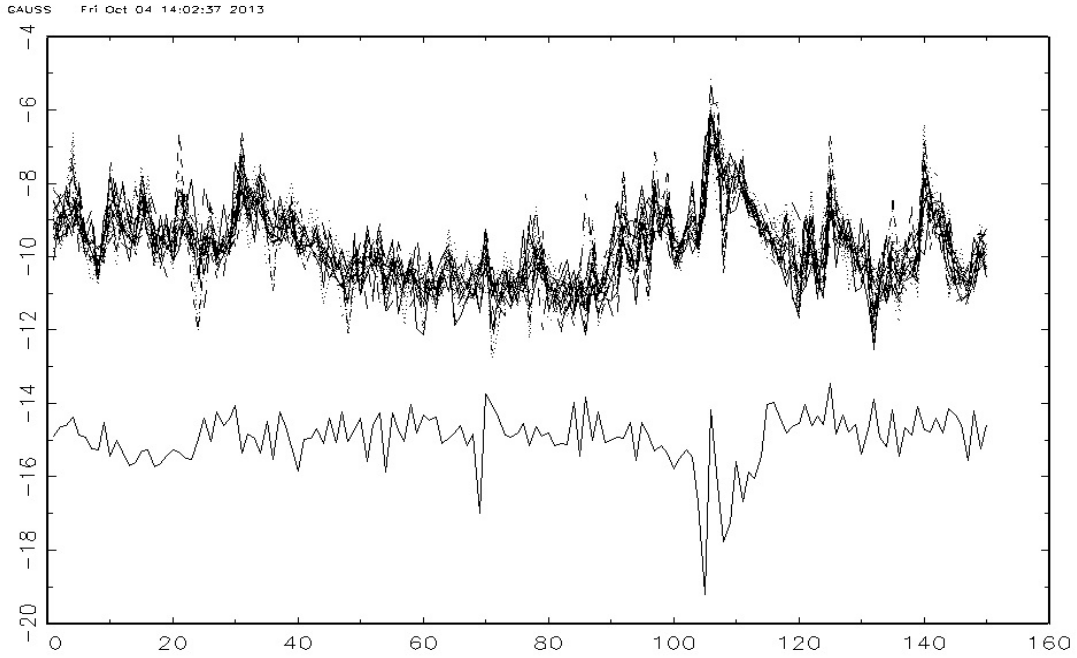


Figure 2:  $ipi$  and  $BV^{(20)}$



Note: This figure shows the monthly growth rate of the U.S. industrial production index (lower line), i.e.,  $ipi$ , and the logarithm of daily bipower variation of the S&P500 stock index (top lines), i.e.,  $BV^{(20)}$ , for the time period from January 2000 to June 2012. The graph for the former is shifted downwards to enable a visual separation of  $ipi$  from the  $bv$ -lines.